Airbnb Listings in Paris, France*

Tam Ly

March 5, 2024

1 Introduction

In this case study we look at Airbnb listings in Paris, France, as of 12 December 2023.

2 Distribution and properties of individual variables

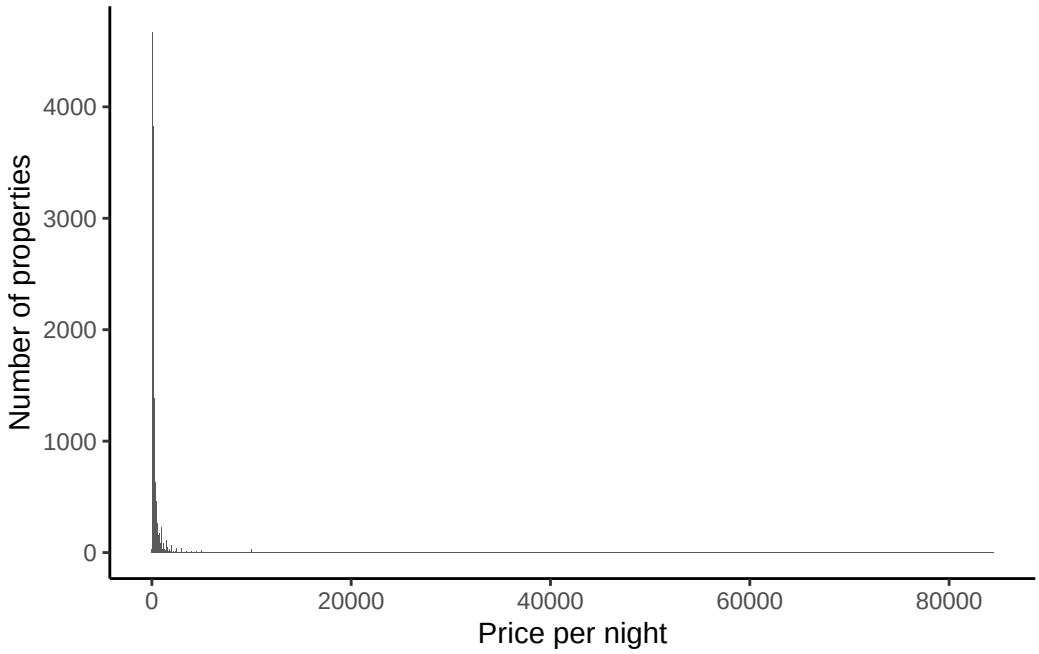
2.1 Distribution of prices of Paris Airbnb rentals in December 2023

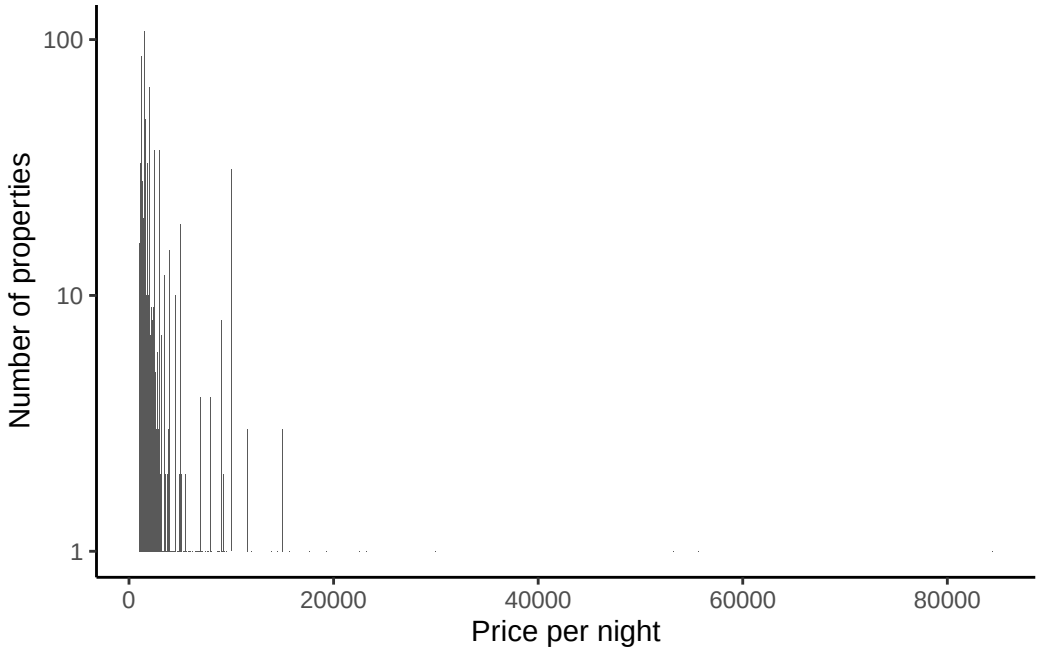
We need to convert the variable from character to numeric:

x
150
146
110
140
180
71

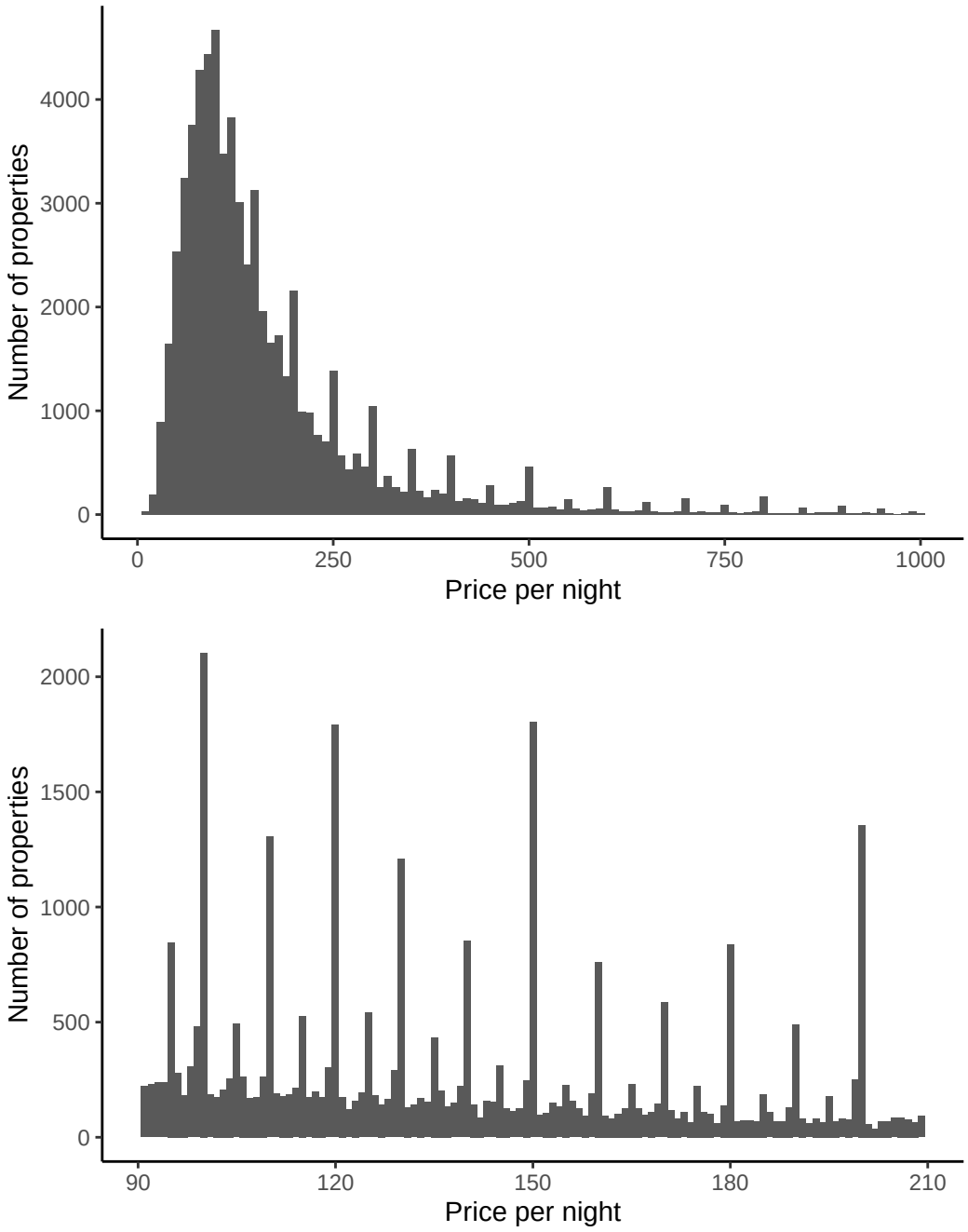
Now we can look at the distribution of prices:

*Code and data are available at: https://github.com/atn-ly/airbnb_paris_eda





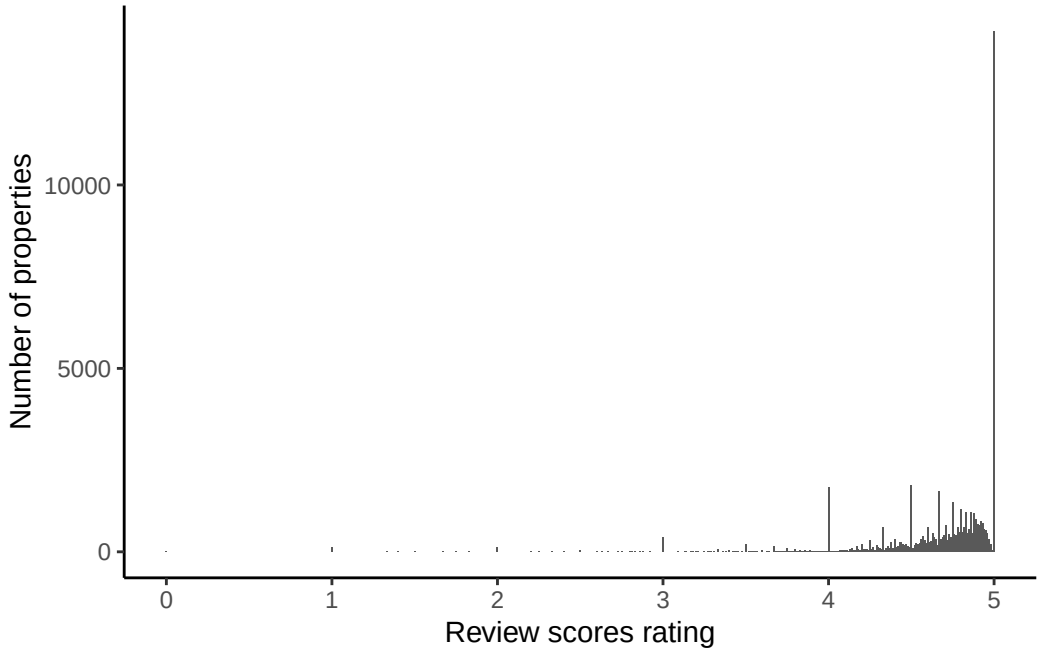
Changing the bins to be smaller:

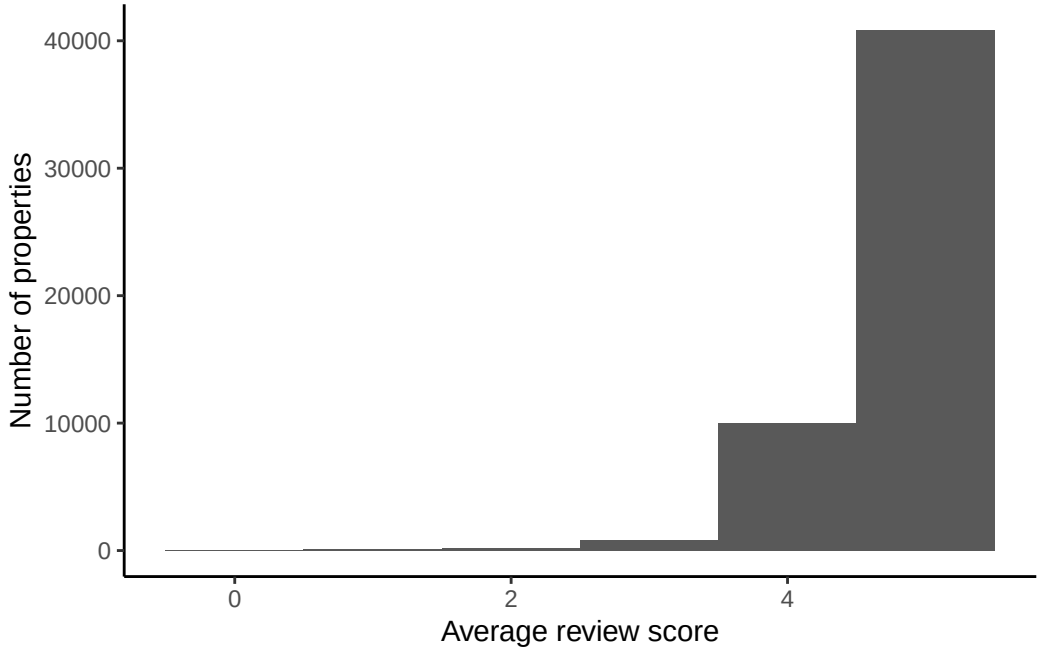


2.2 Distribution of review scores for Paris Airbnb rentals in December 2023

We will remove prices that are more than $999 and remove anyone with a NA for whether they are a superhost.

Now we can look at the distribution of reviews:



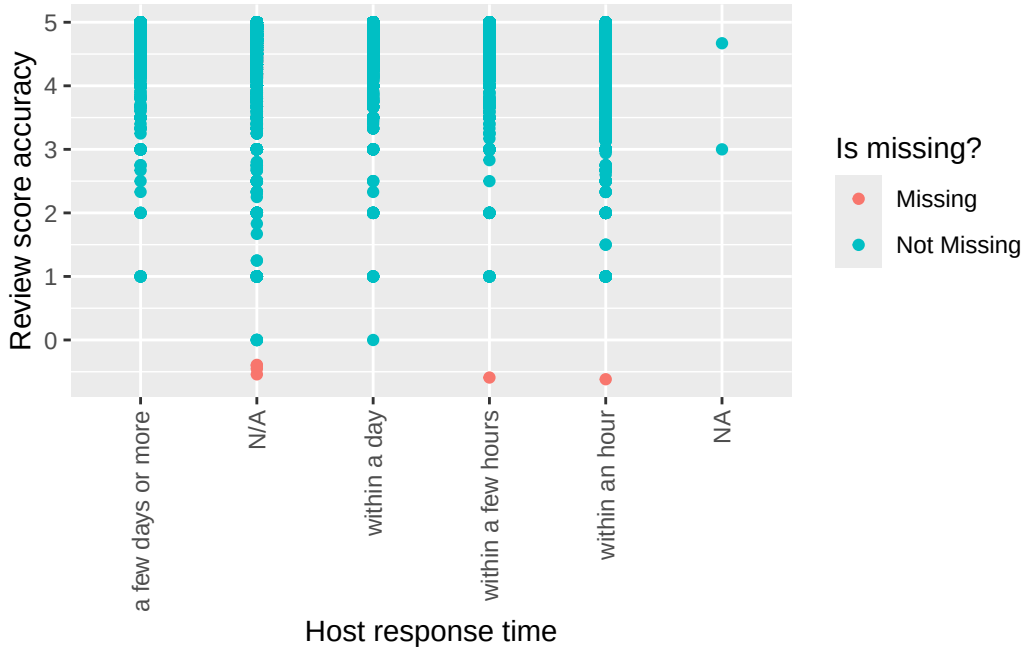


2.3 Missing values in host response time

We see that the variable host_response_time has NAs:

host_response_time	n
N/A	16531
a few days or more	1243
within a day	5297
within a few hours	6811
within an hour	22094
NA	2

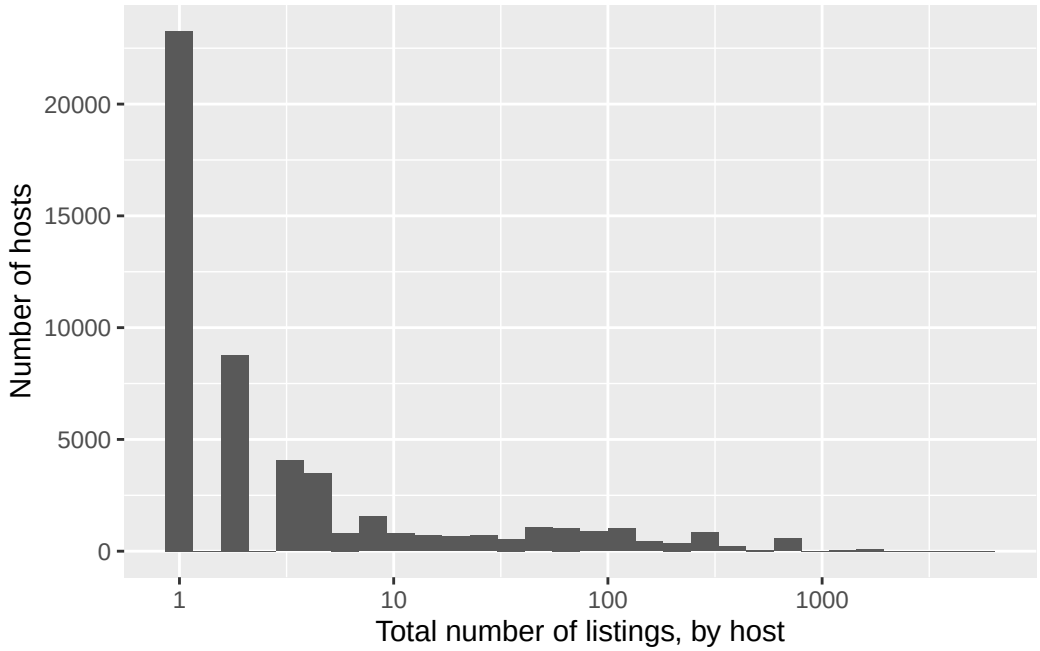
When we include NAs in the graph:



We remove NAs in response time.

```
airbnb_data_selected <-
  airbnb_data_has_reviews |>
  filter(!is.na(host_response_time))
```

2.4 Distribution of the number of properties a host has on Airbnb, for Paris Airbnb rentals in December 2023




```
airbnb_data_selected |>
  filter(host_total_listings_count >= 500) |>
  head()
```

```
# A tibble: 6 x 13
    host_id host_response_time host_is_superhost host_total_listings_count
      <dbl> <chr>              <lgl>                                 <dbl>
1 402191311 N/A                FALSE                                  1679
2  50502817 within an hour     FALSE                                   778
3  50502817 within an hour     FALSE                                   778
4  50502817 within an hour     FALSE                                   778
5  50502817 within an hour     FALSE                                   778
6  50502817 within an hour     FALSE                                   778
# i 9 more variables: neighbourhood_cleansed <chr>, bathrooms <lgl>,
#   bedrooms <dbl>, price <int>, number_of_reviews <dbl>,
#   review_scores_rating <dbl>, review_scores_accuracy <dbl>,
#   review_scores_value <dbl>, host_is_superhost_binary <dbl>
```

We focus on only those with one property for simplicity.

```
airbnb_data_selected <-
  airbnb_data_selected |>
  add_count(host_id) |>
  filter(n == 1) |>
  select(-n)
```

3 Relationships between variables

3.1