Airbnb Listings in Paris, France*

Tam Ly

March 5, 2024

Table of contents

1 Introduction **1**

2 Distribution and properties of individual variables **2**
2.1 Distribution of prices of Paris Airbnb rentals in December 2023 2
2.2 Distribution of review scores for Paris Airbnb rentals in December 2023 4
2.3 Missing values in host response time 5
2.4 Distribution of the number of properties a host has on Airbnb, for Paris Airbnb rentals in December 2023 7

3 Relationships between variables **8**
3.1 Relationship between price and review and whether a host is a superhost, for Paris Airbnb rentals in December 2023 8
3.2 Model explaining whether a host is a superhost based on their response time . 8

References **9**

1 Introduction

In this case study we look at Airbnb listings in Paris, France, as of 12 December 2023. We use the same data from the textbook (Alexander 2023) in R (R Core Team 2022) and the packages `tidyverse` (Wickham et al. 2019), `knitr` (Xie 2023), `arrow` (Richardson et al. 2024), `naniar` (Tierney and Cook 2023), and `modelsummary` (Arel-Bundock 2022).

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

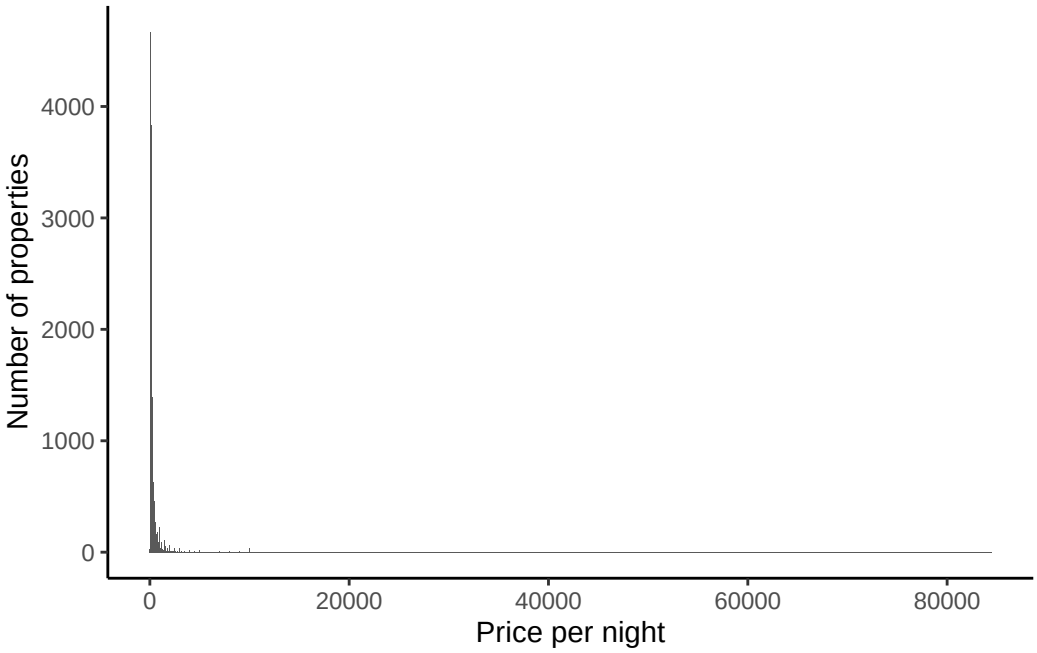
2 Distribution and properties of individual variables

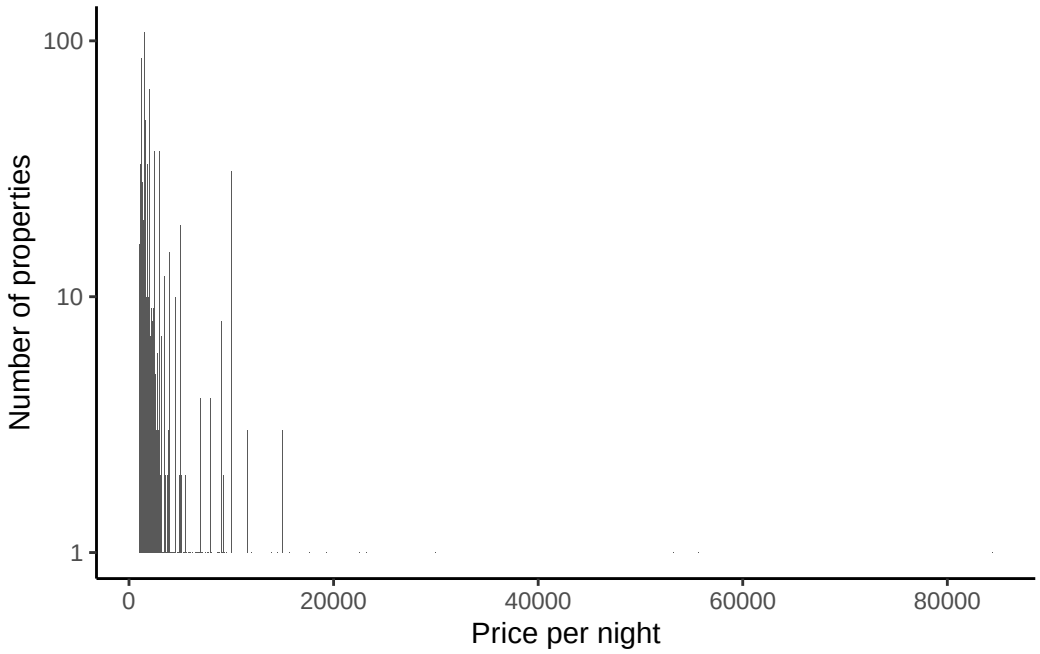
2.1 Distribution of prices of Paris Airbnb rentals in December 2023

We need to convert the variable from character to numeric:

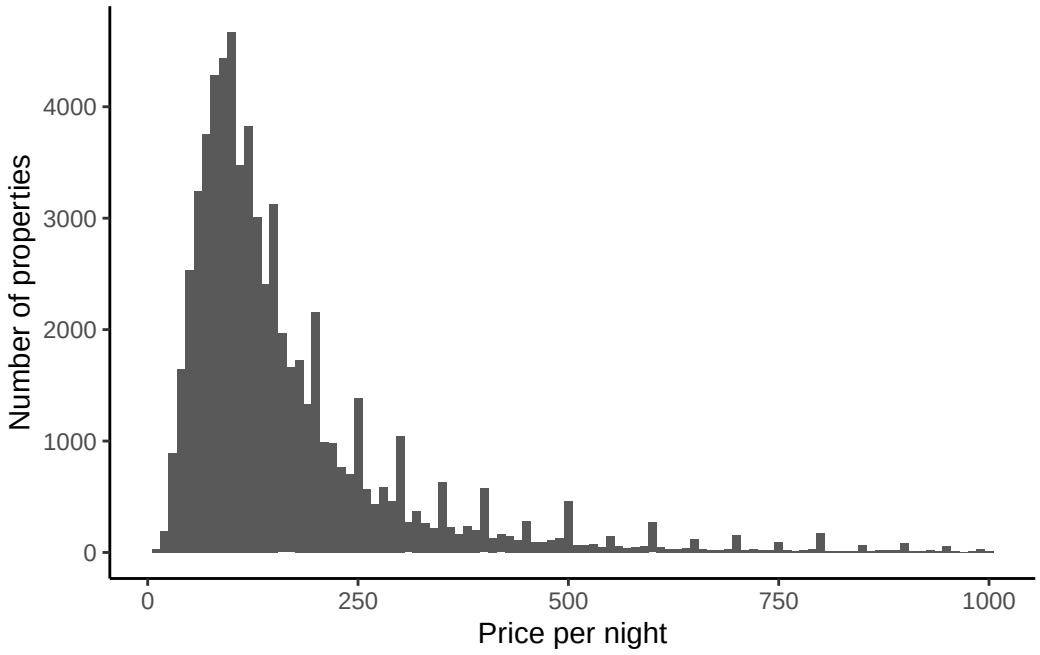
x
150
146
110
140
180
71

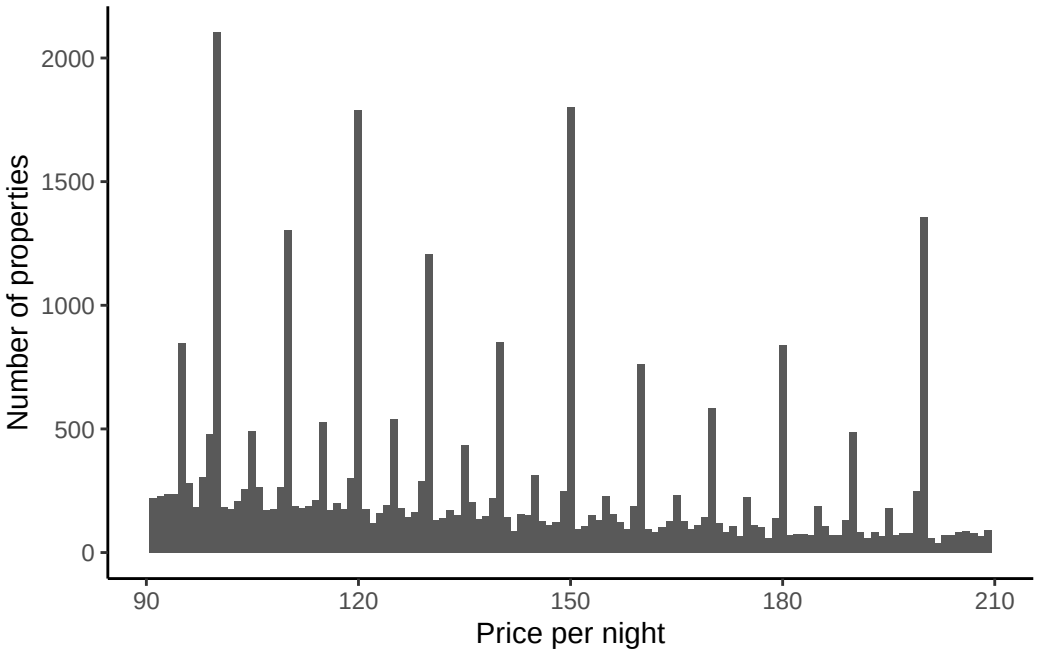
Now we can look at the distribution of prices:





Changing the bins to be smaller:

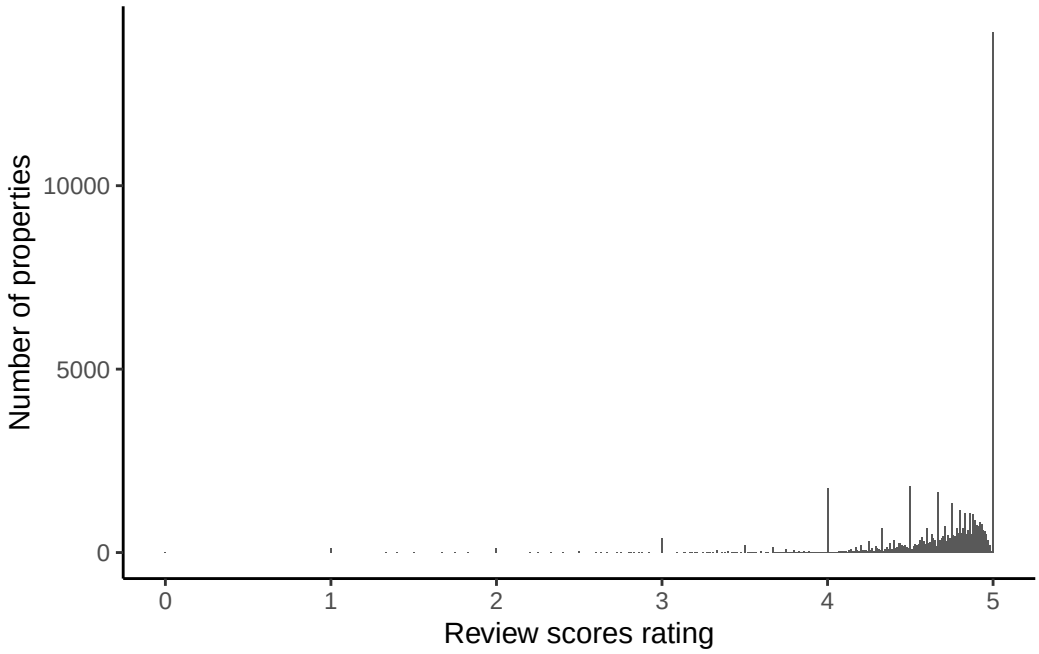


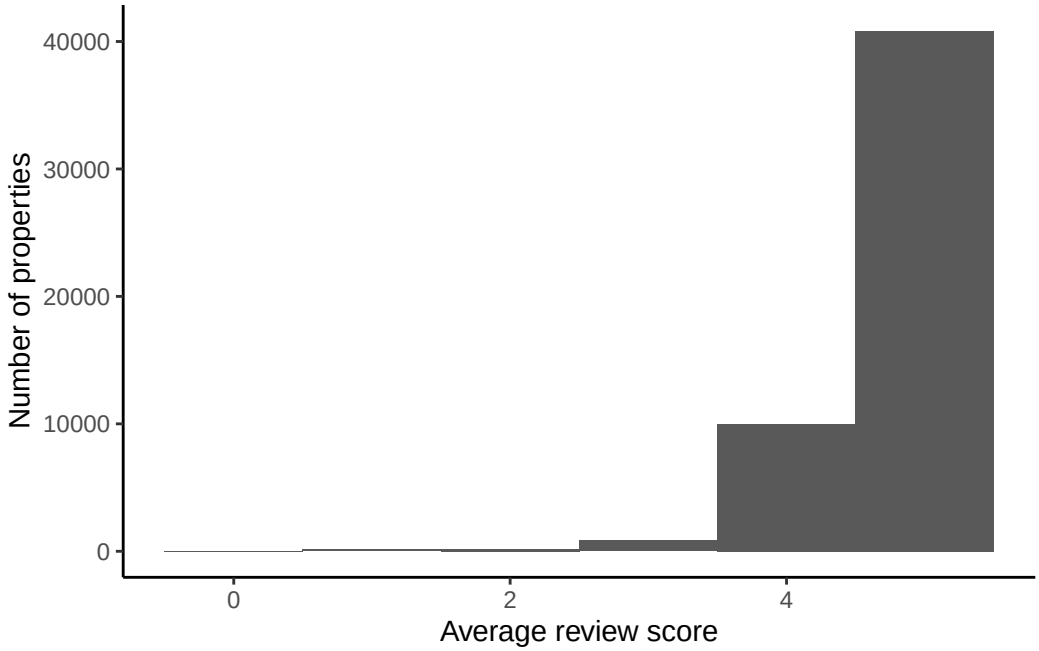


2.2 Distribution of review scores for Paris Airbnb rentals in December 2023

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:



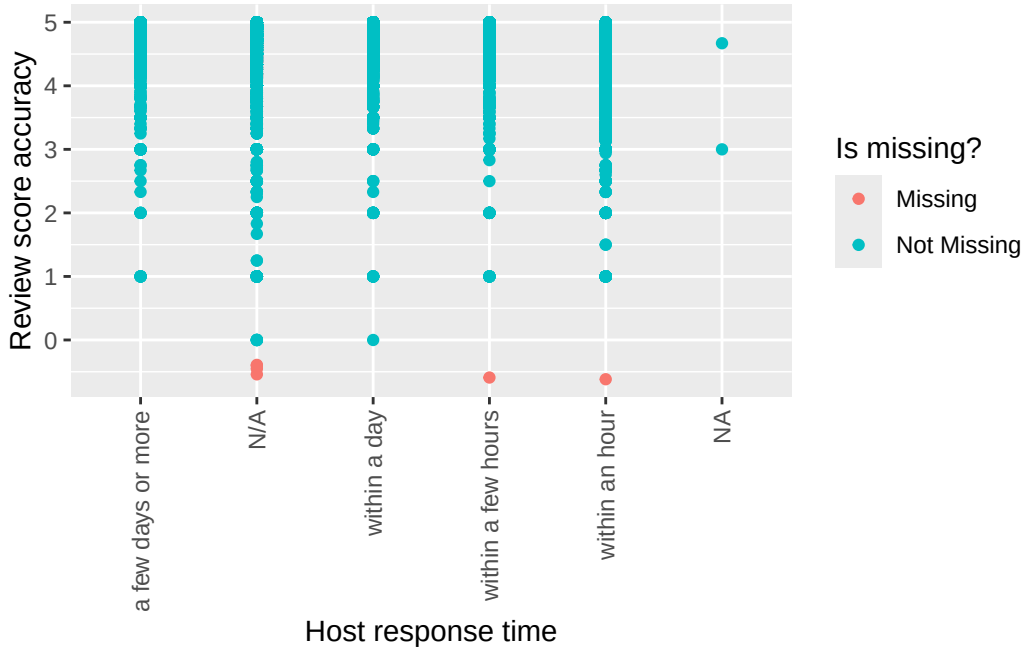


2.3 Missing values in host response time

We see that the variable host_response_time has NAs:

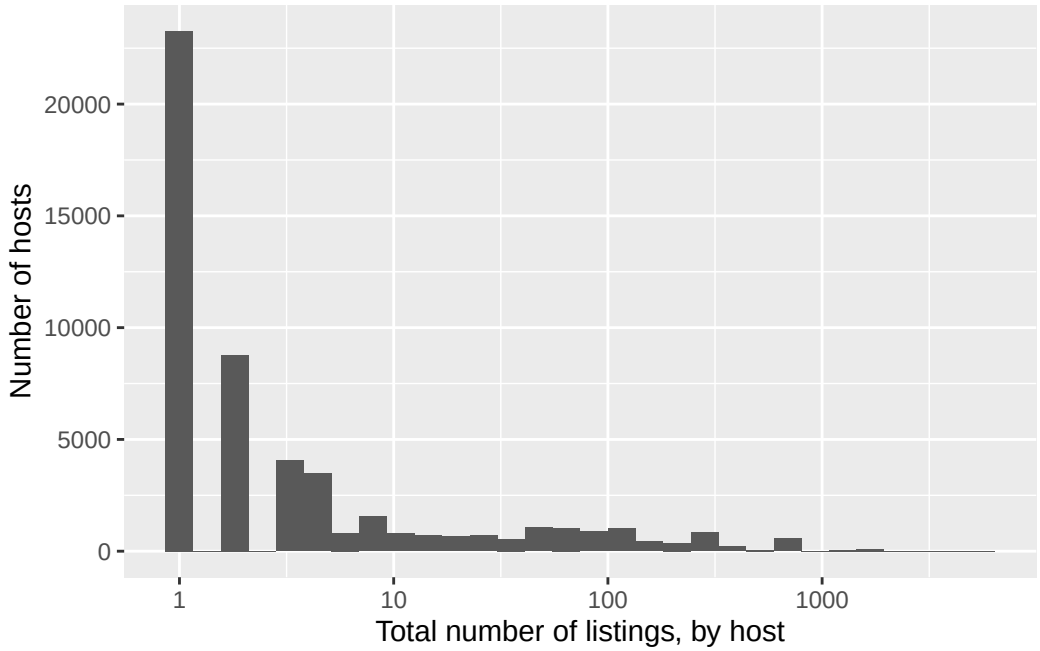
host_response_time	n
N/A	16531
a few days or more	1243
within a day	5297
within a few hours	6811
within an hour	22094
NA	2

When we include NAs in the graph:



We remove NAs in response time.

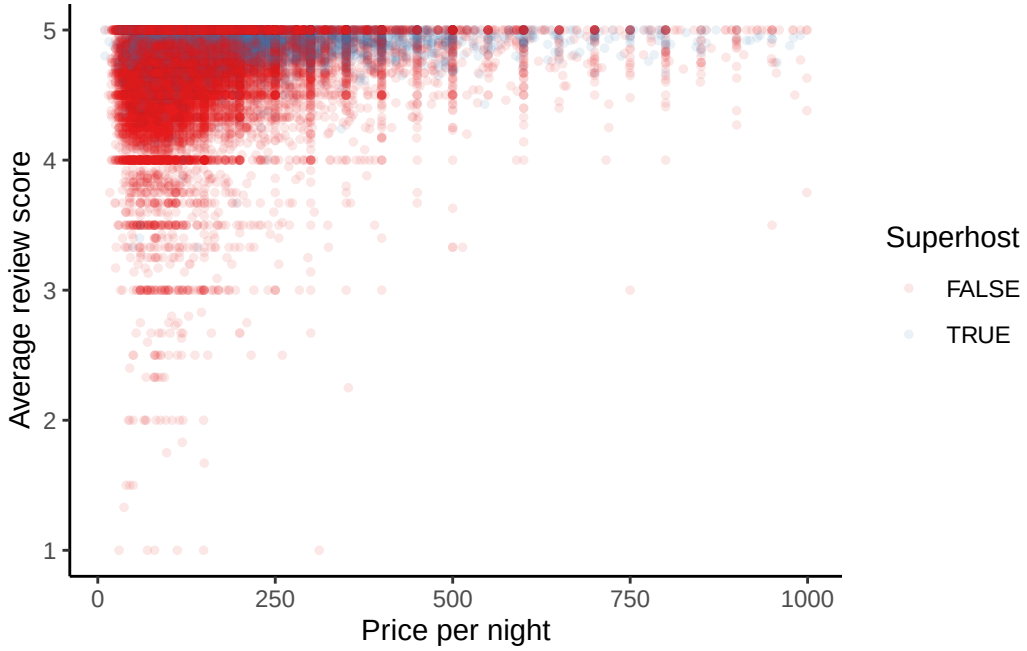
2.4 Distribution of the number of properties a host has on Airbnb, for Paris Airbnb rentals in December 2023



We focus on only those with one property for simplicity.

3 Relationships between variables

3.1 Relationship between price and review and whether a host is a superhost, for Paris Airbnb rentals in December 2023



3.2 Model explaining whether a host is a superhost based on their response time

	(1)
(Intercept)	−16.354
	(0.448)
host_response_timeN/A	0.877
	(0.211)
host_response_timewithin a day	2.019
	(0.211)
host_response_timewithin a few hours	2.695
	(0.210)
host_response_timewithin an hour	2.973
	(0.209)
review_scores_rating	2.643
	(0.082)
Num.Obs.	35 652
AIC	29 972.7
BIC	30 023.6
Log.Lik.	−14 980.357
RMSE	0.37

References

Alexander, Rohan. 2023. *Telling Stories with Data*. Chapman; Hall/CRC. https://tellingstorieswithdata.com.

Arel-Bundock, Vincent. 2022. "modelsummary: Data and Model Summaries in R." *Journal of Statistical Software* 103 (1): 1–23. https://doi.org/10.18637/jss.v103.i01.

R Core Team. 2022. *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing. https://www.R-project.org/.

Richardson, Neal, Ian Cook, Nic Crane, Dewey Dunnington, Romain François, Jonathan Keane, Dragoș Moldovan-Grünfeld, Jeroen Ooms, Jacob Wujciak-Jens, and Apache Arrow. 2024. *Arrow: Integration to 'Apache' 'Arrow'*. https://github.com/apache/arrow/.

Tierney, Nicholas, and Dianne Cook. 2023. "Expanding Tidy Data Principles to Facilitate Missing Data Exploration, Visualization and Assessment of Imputations." *Journal of Statistical Software* 105 (7): 1–31. https://doi.org/10.18637/jss.v105.i07.

Wickham, Hadley, Mara Averick, Jennifer Bryan, Winston Chang, Lucy D'Agostino McGowan, Romain François, Garrett Grolemund, et al. 2019. "Welcome to the tidyverse." *Journal of Open Source Software* 4 (43): 1686. https://doi.org/10.21105/joss.01686.

Xie, Yihui. 2023. *Knitr: A General-Purpose Package for Dynamic Report Generation in r*. https://yihui.org/knitr/.