#datosturismobcn2018
28 y 29 de septiembre 2018

Toda la info y links en:
montera34/bcn2018

Simposio: viernes mañana.
- Ronda presentaciones
- Ponencias
- Debate

Taller: viernes tarde y sábado todo el día

Documentación

Documento colaborativo ¡ayuda a documentar!
Wiki: wiki.montera34.com

¿Cómo analizar algo que no quieren que sepas?

Los problemas de analizar Airbnb y otras plataformas de viviendas de uso turístico.

Montera34
Pablo Rey Mazón
Alfonso Sánchez Uzábal

Barcelona, 28 septiembre 2018
Taller de datos sobre alojamientos turísticos

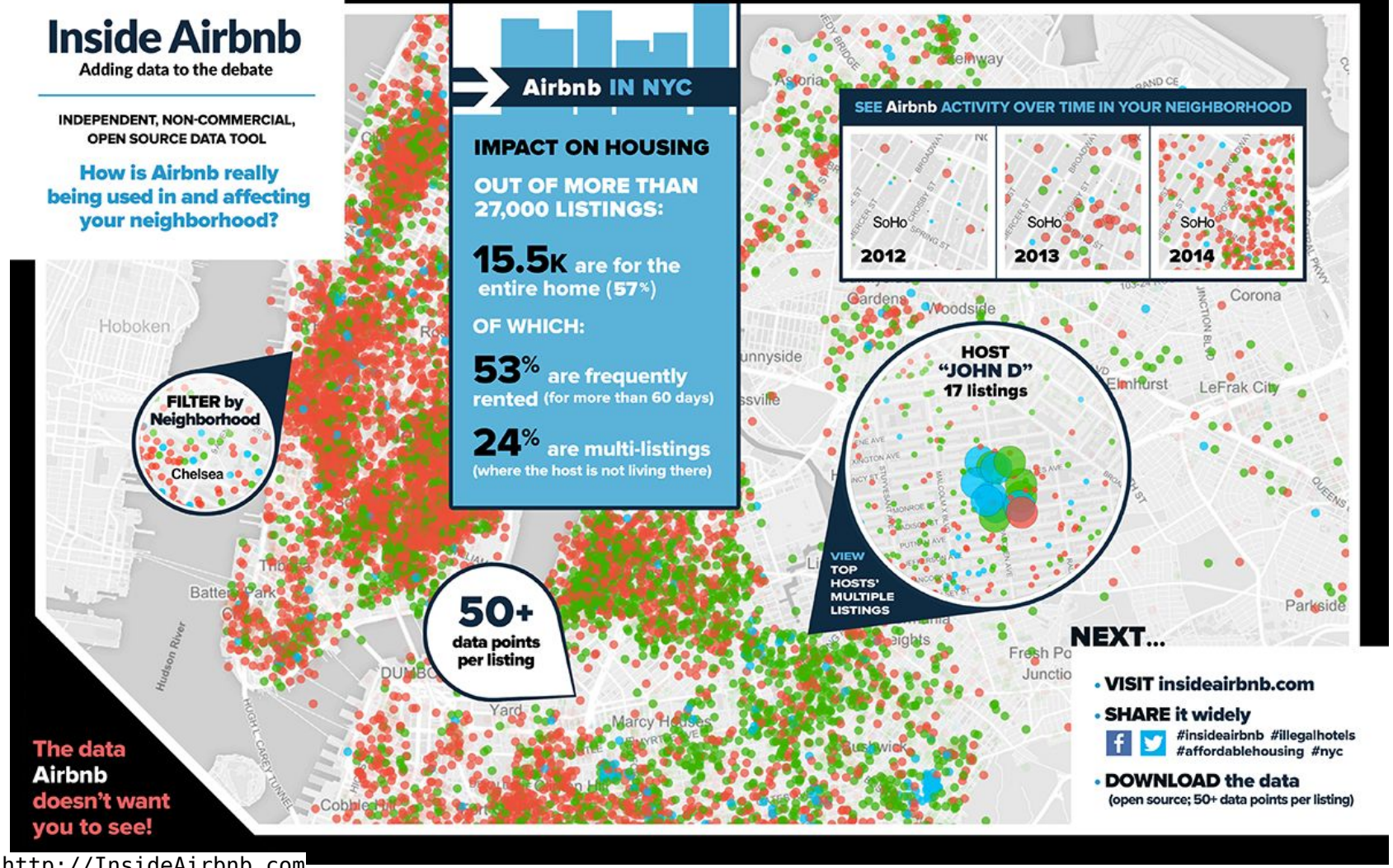


Imagen: http://InsideAirbnb.com

Conocer las fuentes de datos y su metodología de obtención
Desconfiemos de los datos
Pongamos en duda las visualizaciones

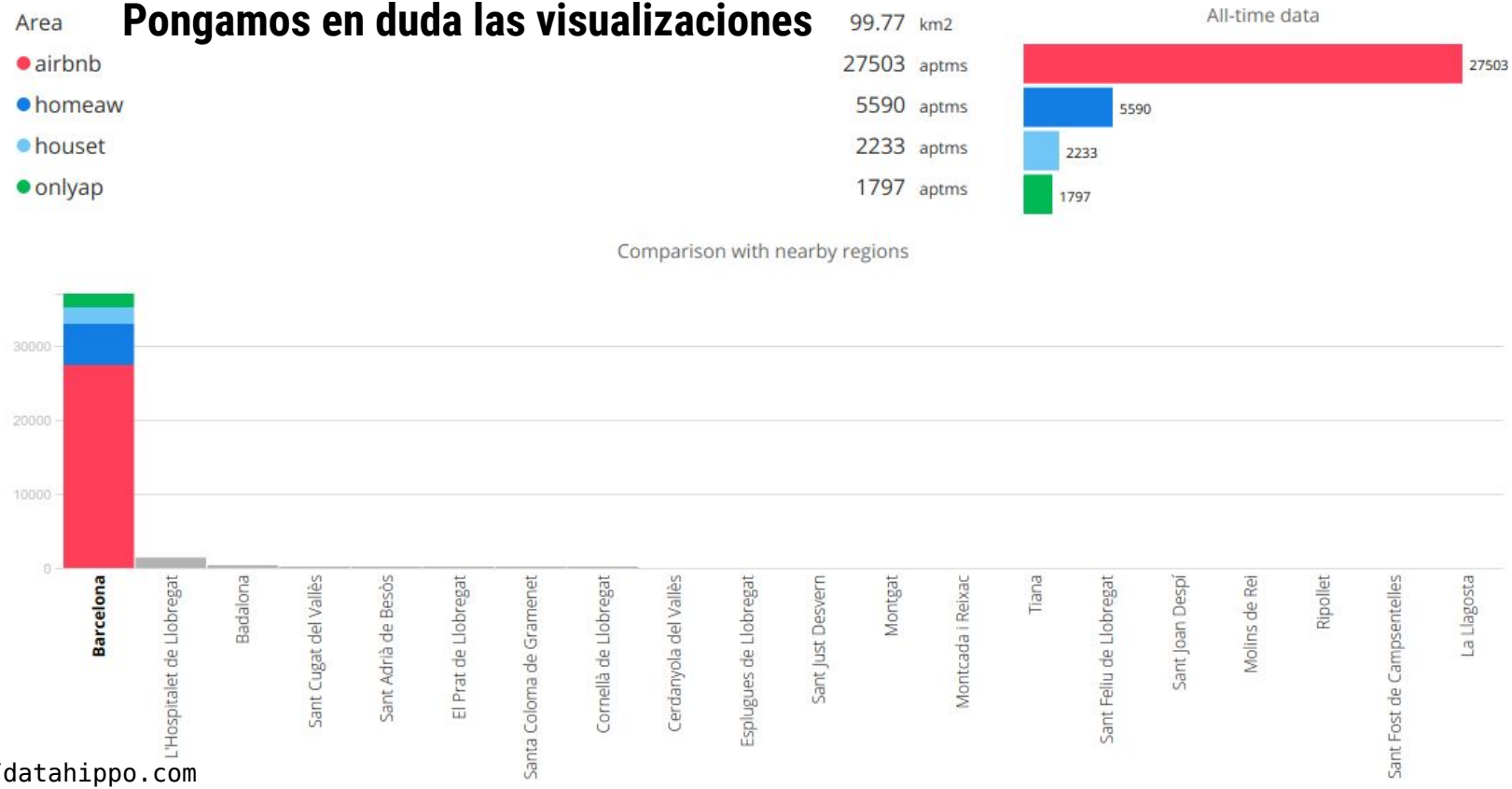
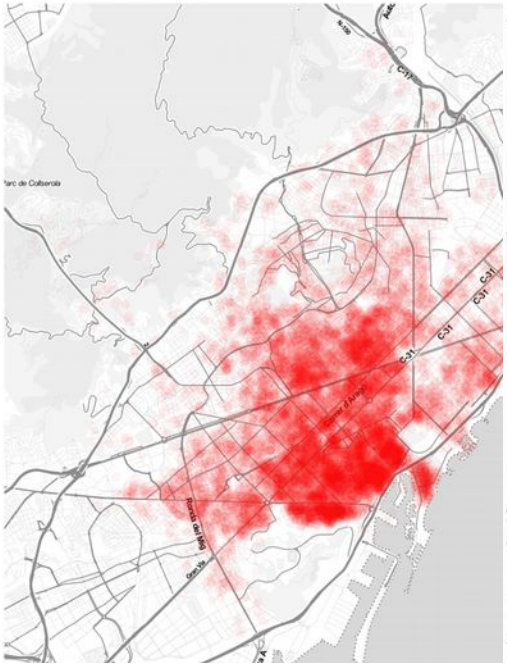


Imagen: http://datahippo.com

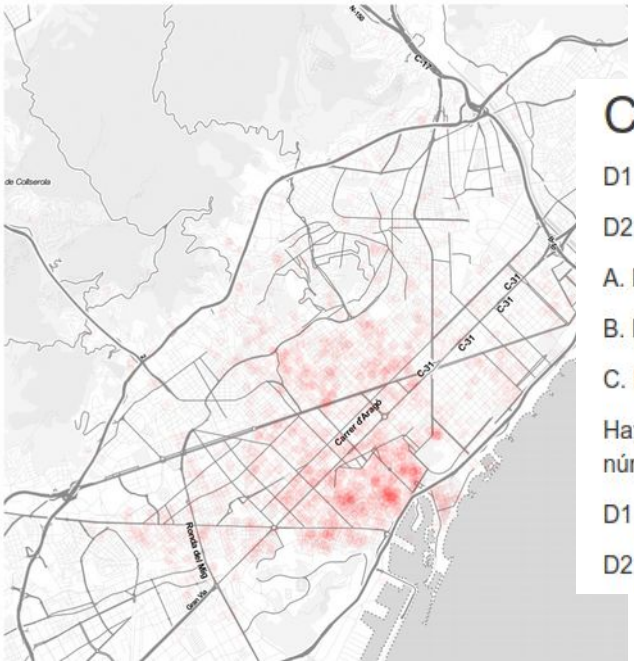
Comparar bases de datos

D1 Inside Airbnb contiene 19.261 anuncios (18 agosto 2018)
D2 Inside Airbnb contiene 19.200 anuncios (19 septiembre 2018)

DB1: data set 1: IA



anuncios D1 que no están en data set D2



Comparando datasets

D1 tiene **19261** anuncios.

D2 tiene **19200** anuncios.

A. D1 e D2 tienen **17258** anuncios comunes.

B. D1 tiene **2003** anuncios que no están en la D2.

C. D2 tiene **1942** anuncios que no están en la D1.

Hay, por tanto **21203 anuncios** (A + B + C) diferentes en total entre las d
número de valores de id únicos entre d1 y d2.

D1 tiene **90.8%** de los anuncios combinados de las dos bases de datos.

D2 tiene **90.6%** de los anuncios combinados de las dos bases de datos.

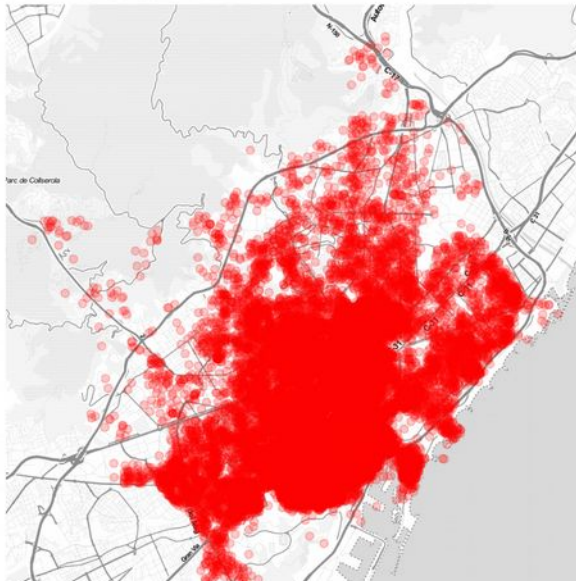
Imagen: https://github.com/montera34/airbnb.barcelona

Comparar bases de datos

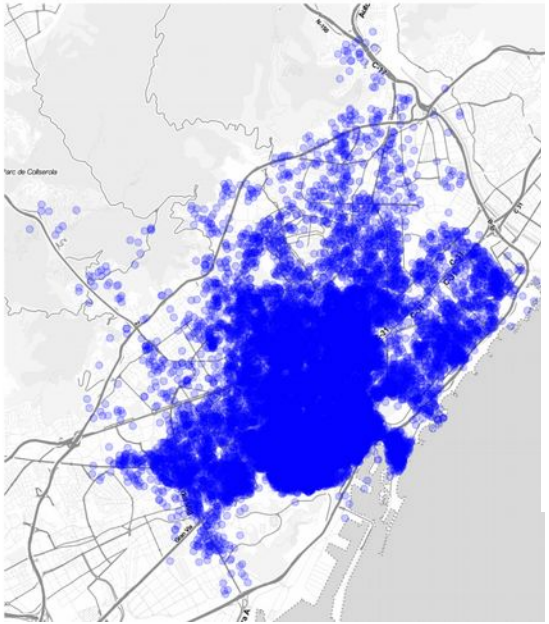
D1 Datahippo 24.029 anuncios (19 junio 2018)

D2 Inside Airbnb 19.261 anuncios (18 agosto 2018)

DB1: data set 1



DB2: data set 2



Comparando datasets

DH tiene **24029** anuncios.

IA tiene **19261** anuncios.

A. DH e IA tienen **13207** anuncios comunes.

B. DH tiene **10822** anuncios que no están en la IA.

C. IA tiene **6054** anuncios que no están en la DH.

Hay, por tanto **30083 anuncios** (A + B + C) diferentes en total entre las dos bases d
número de valores de id únicos entre DH e IA.

DH tiene **79.9%** de los anuncios combinados de las dos bases de datos.

AI tiene **64%** de los anuncios combinados de las dos bases de datos.

Imagen: https://github.com/montera34/airbnb.barcelona

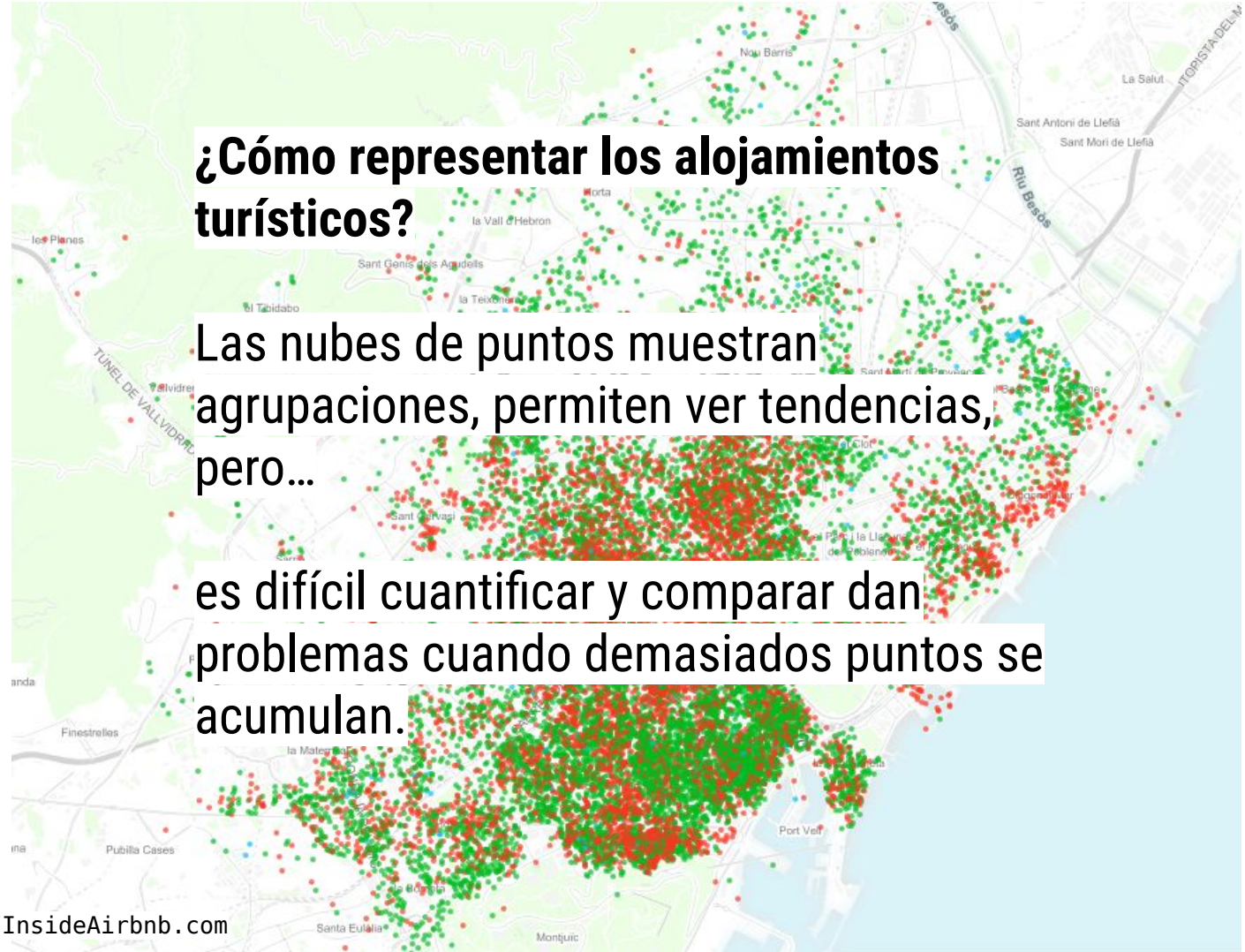


Imagen: http://InsideAirbnb.com

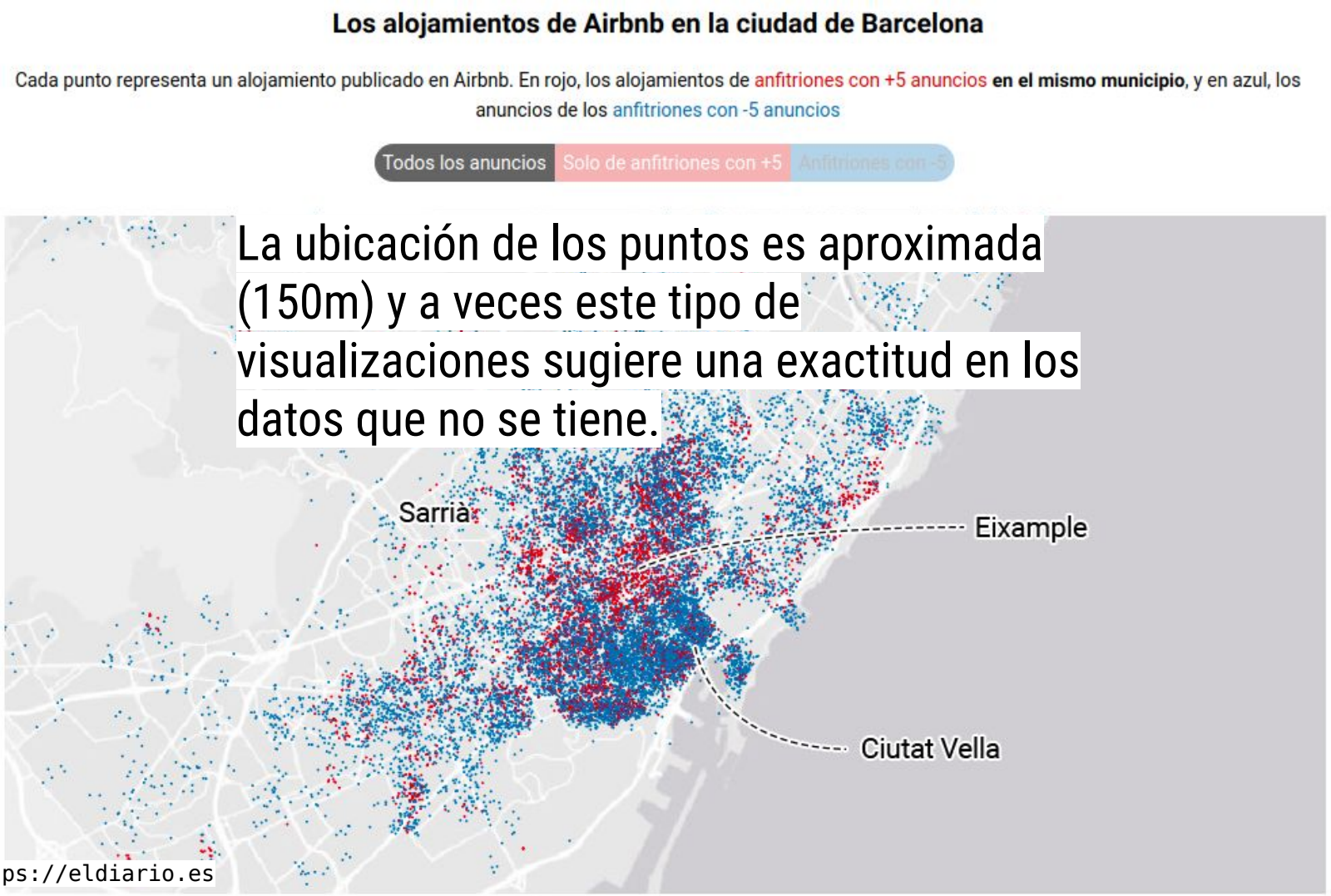
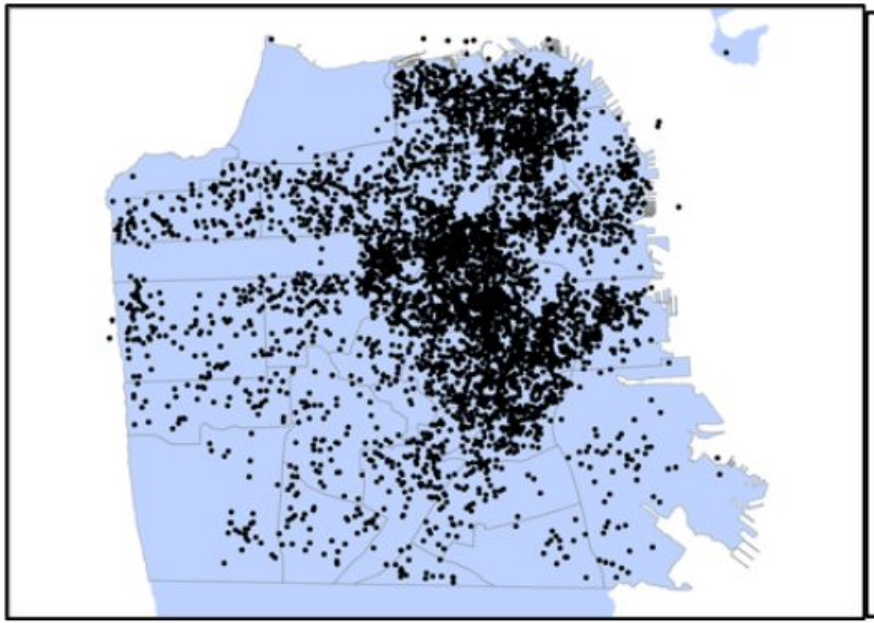
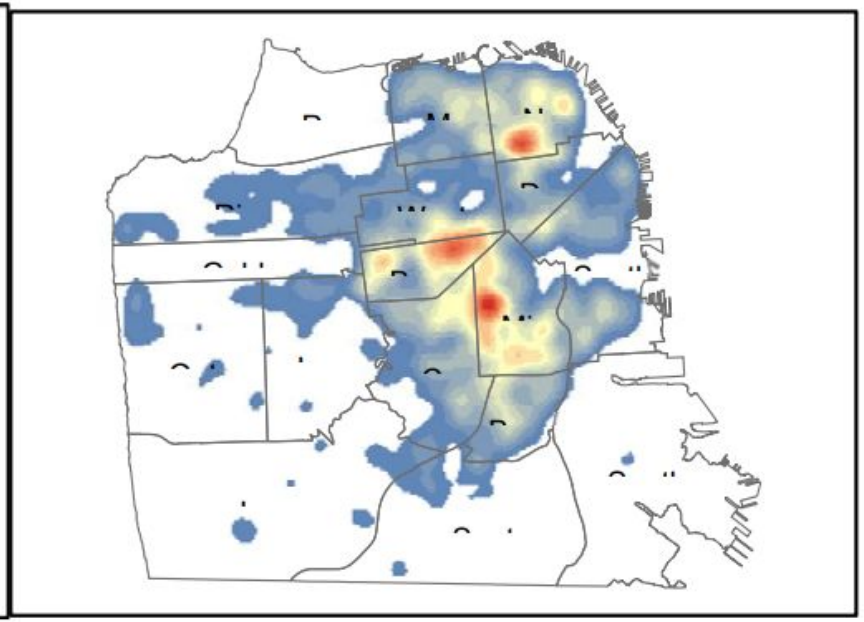


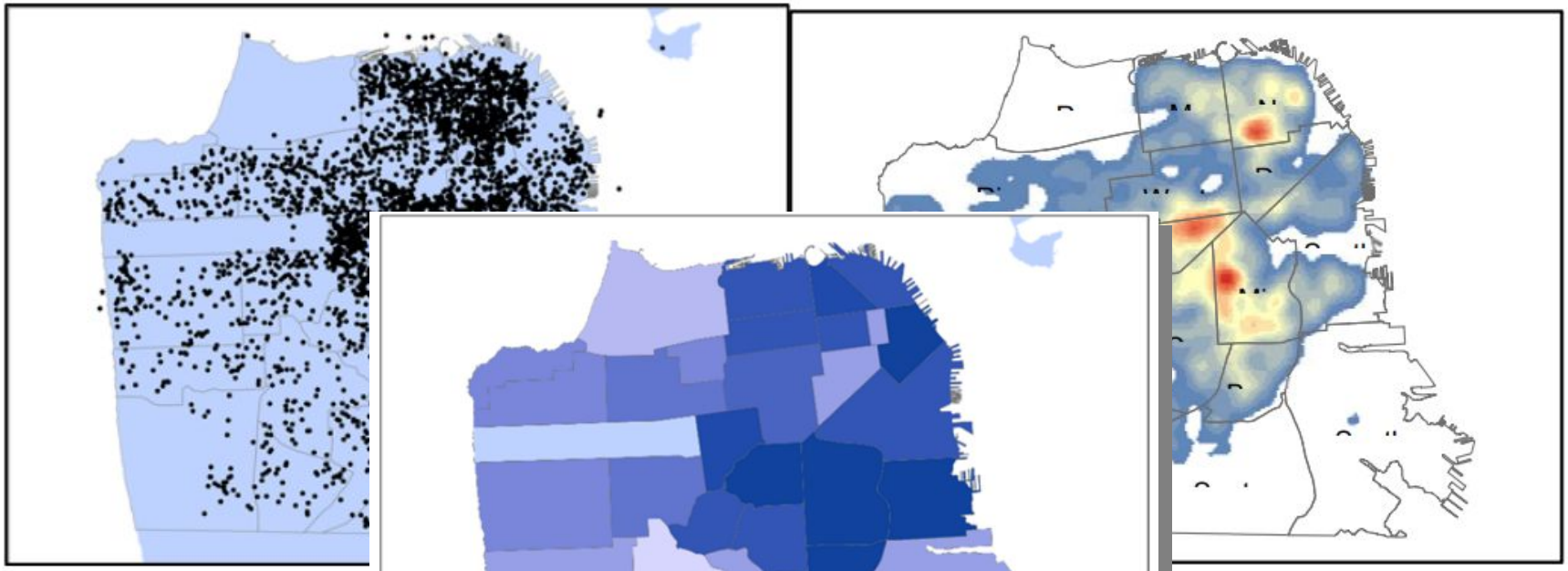
Imagen: https://eldiario.es





As the maps above show, Airbnb units appear to be concentrated in the northern and eastern neighborhoods of the city. However, this map merely shows where units are located and does not take into account the fact that those parts of the city have a higher density of all housing units. It is necessary to control for the density of housing units in order to identify if there are areas of the city with higher concentrations of Airbnb units. The map below controls for neighborhood density by dividing the number of Airbnb units by the total number of housing units for each neighborhood.

Imagen: http:// http://commissions.sfplanning.org/cpcpackets/2014-001033PCA.pdf



As the maps above shov rthern and eastern neighborhoods of the ci re located and does not take into account the fact th housing units. It is necessary to control for here are areas of the city with higher concentrations of Airbnb units. The map below controls for neighborhood density by dividing the number of Airbnb units by the total number of housing units for each neighborhood.

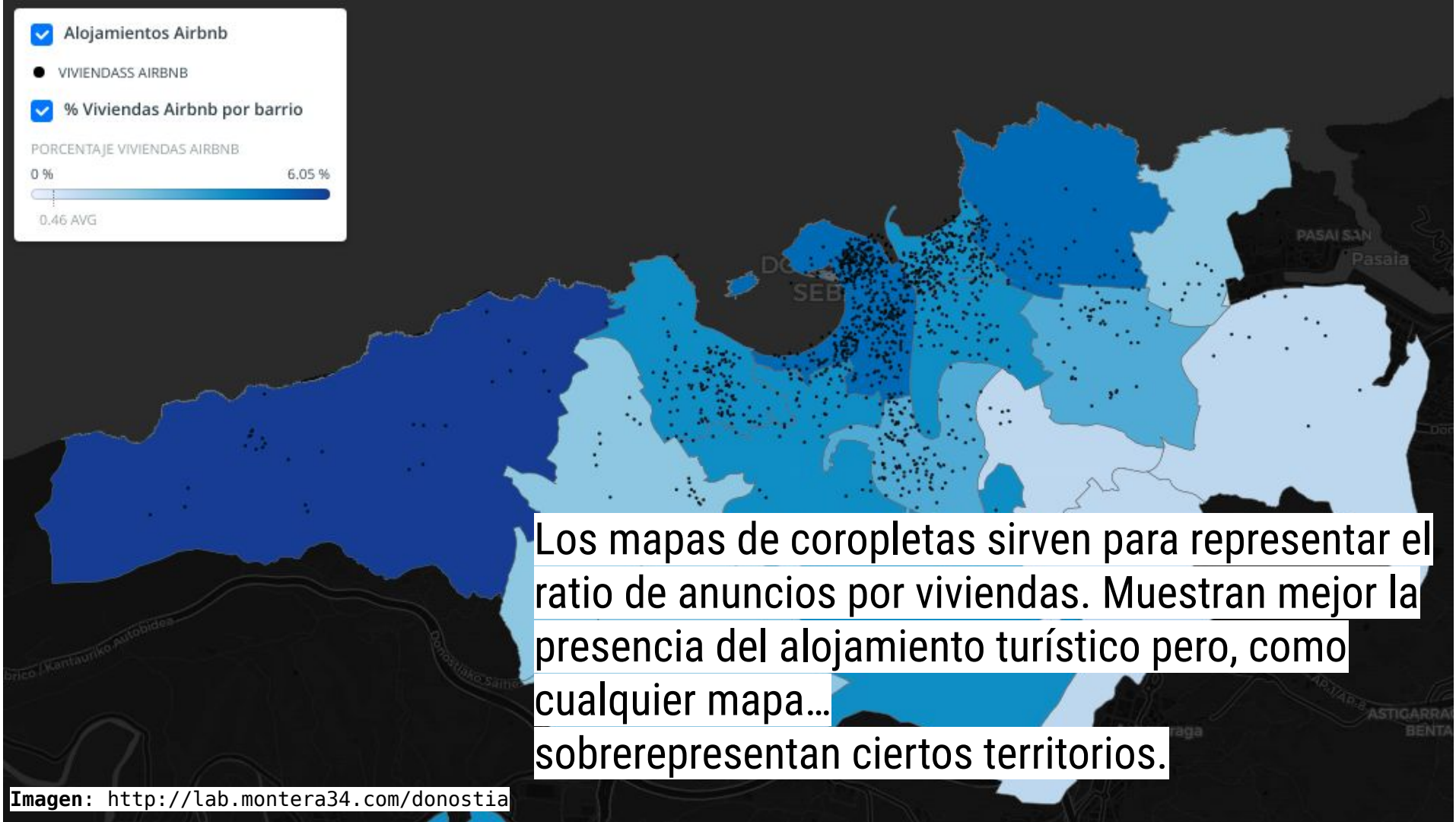
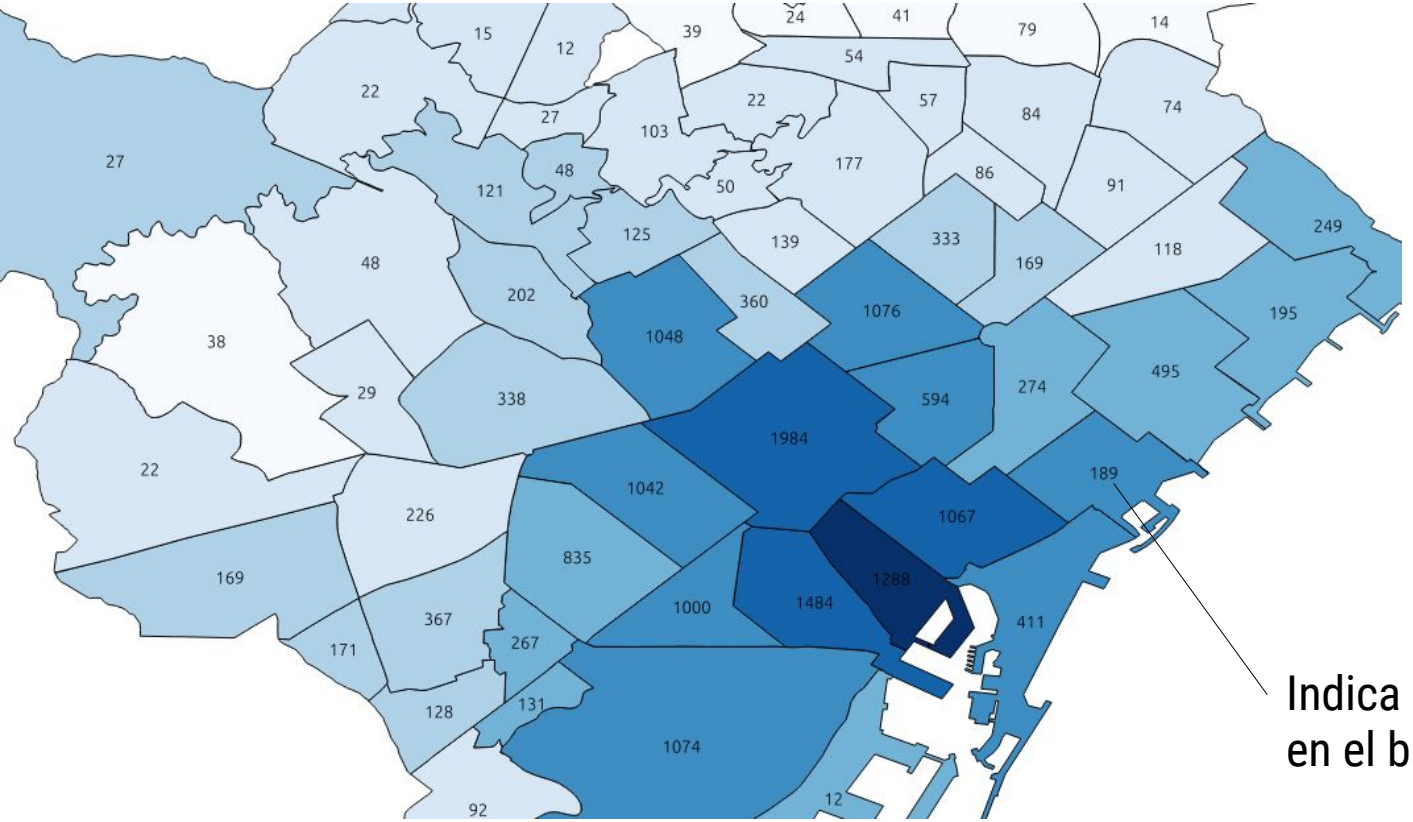


Imagen: http://lab.montera34.com/donostia



Ratio anuncios Airbnb por cada 100 viviendas

- 0.0 - 0.0
- 0.0 - 1.0
- 1.0 - 2.0
- 2.0 - 4.0
- 4.0 - 7.0
- 7.0 - 11.0
- 11.0 - 20.0

Indica el número de anuncios en el barrio

No siempre necesitamos mapas.

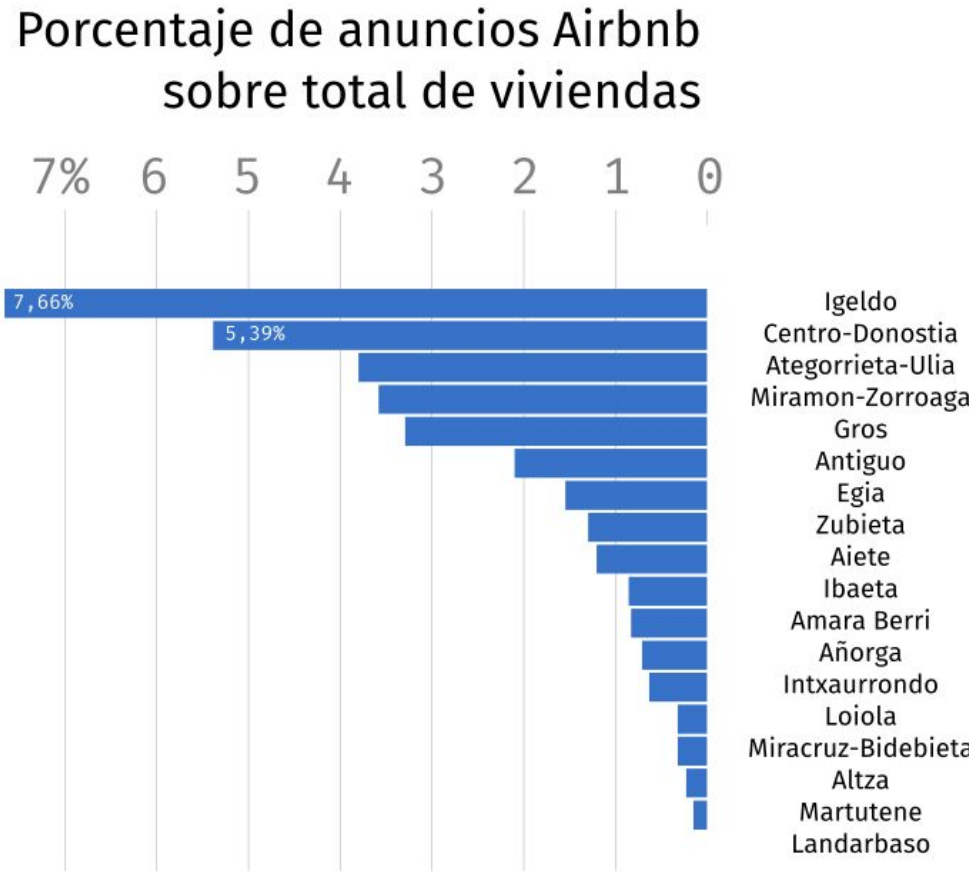


Imagen: http://lab.montera34.com/donostia

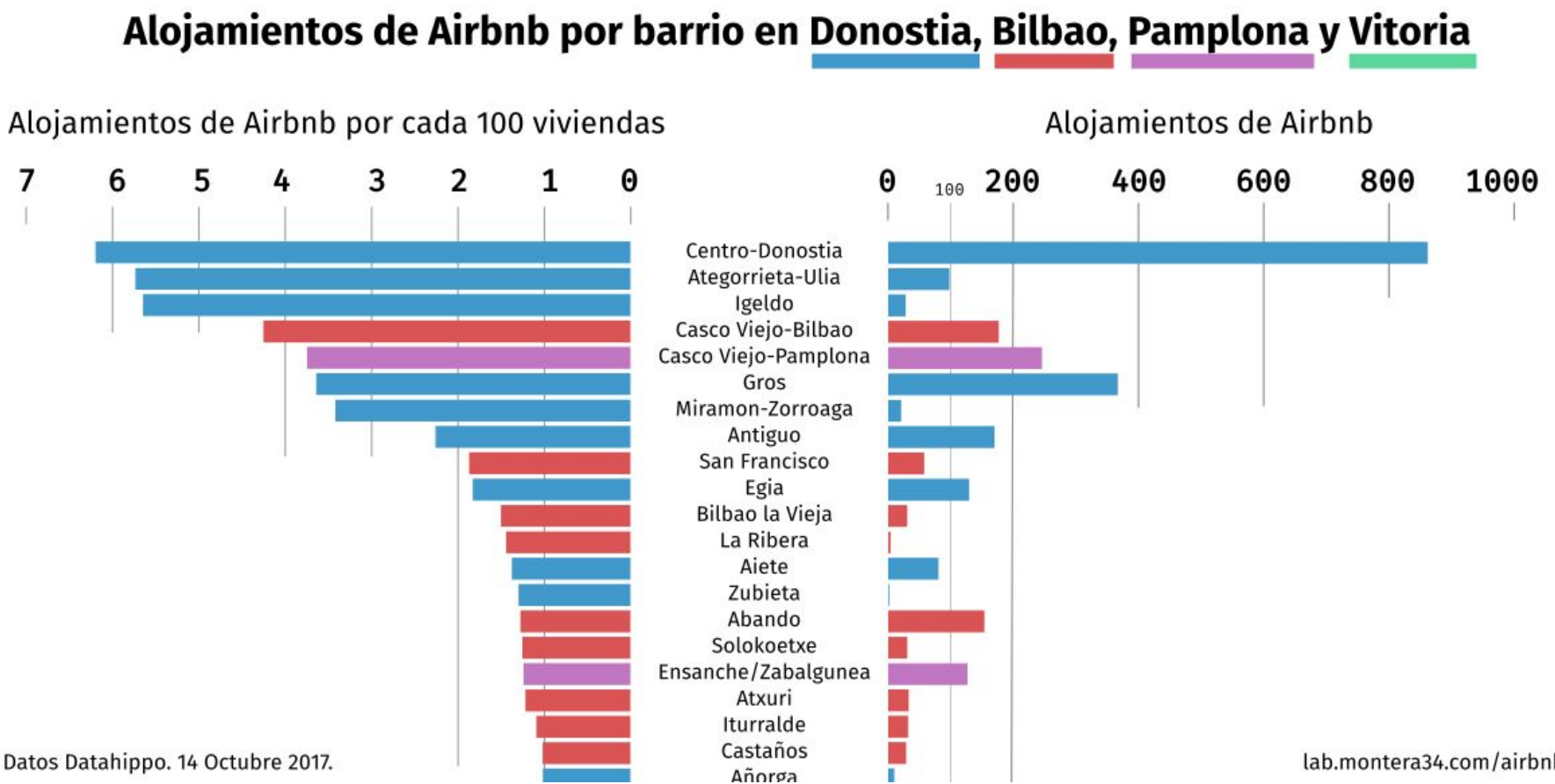


Imagen: http://lab.montera34.com/barrios

¿Estamos contando bien?
¿Qué alojamientos debemos contar?
¿Qué es un alojamiento activo?
¿Cuándo un alojamiento se convierte en un hotel?

Podemos analizarlo:
- Pasado: según número de reviews
- Presente: ¿está la URL activa?
- Futuro: fechas disponibles para ser alquilado
- Según número medio de reviews por mes.

¿Activo para qué?

Airbnb en Donostia

Fecha	Descargado en Abril 2018		Descargado en septiembre 2018 (última fecha de *revised*: 2018-06-04)				
Fuente	Datahippo		Datahippo				
			total	disponible	existe	Review posterior	Si el 72% de usuarios dejara review...
	Listado total	Tenían fechas disponibles a fecha 18-04-01	Listado total	Tenían fechas disponibles a fecha 18-06-04	Existe la URL 2018-09-11	Ultima review posterior a 2018-03-31	
Habitaciones	471	71	557	355	339	234	325
Habitaciones compartidas	8	2	10	6	6	3	4
Viviendas	1.390	191	1.639	1.076	958	763	1.059
Total	1.869	264	**2.206** 100%	**1.437** 65%	**1.303** 59%	**1.000** 45%	**1.389** 62%

	Listado total	Tenían fechas disponibles a fecha junio 2018	Existe la URL septiembre 2018	Última review posterior a 2018-04-01
Donostia anuncios	**2.206**	**1.437**	**1.303**	**1.000**
	100%	65%	59%	45%
Bilbao anuncios	**1.301**	**897**	**774**	**611**
	100%	69%	59%	47%

El objetivo es poder responder a:

"Muestra los alojamientos que están habitualmente disponibles en esta ciudad que sean de vivienda completa, que hayan recibido una “review” en los últimos 6 meses Y se alquilen frecuentemente Y el anfitrión gestione también otros alojamientos."

Adaptado de http://insideairbnb.com/about.html

La importancia de filtrar

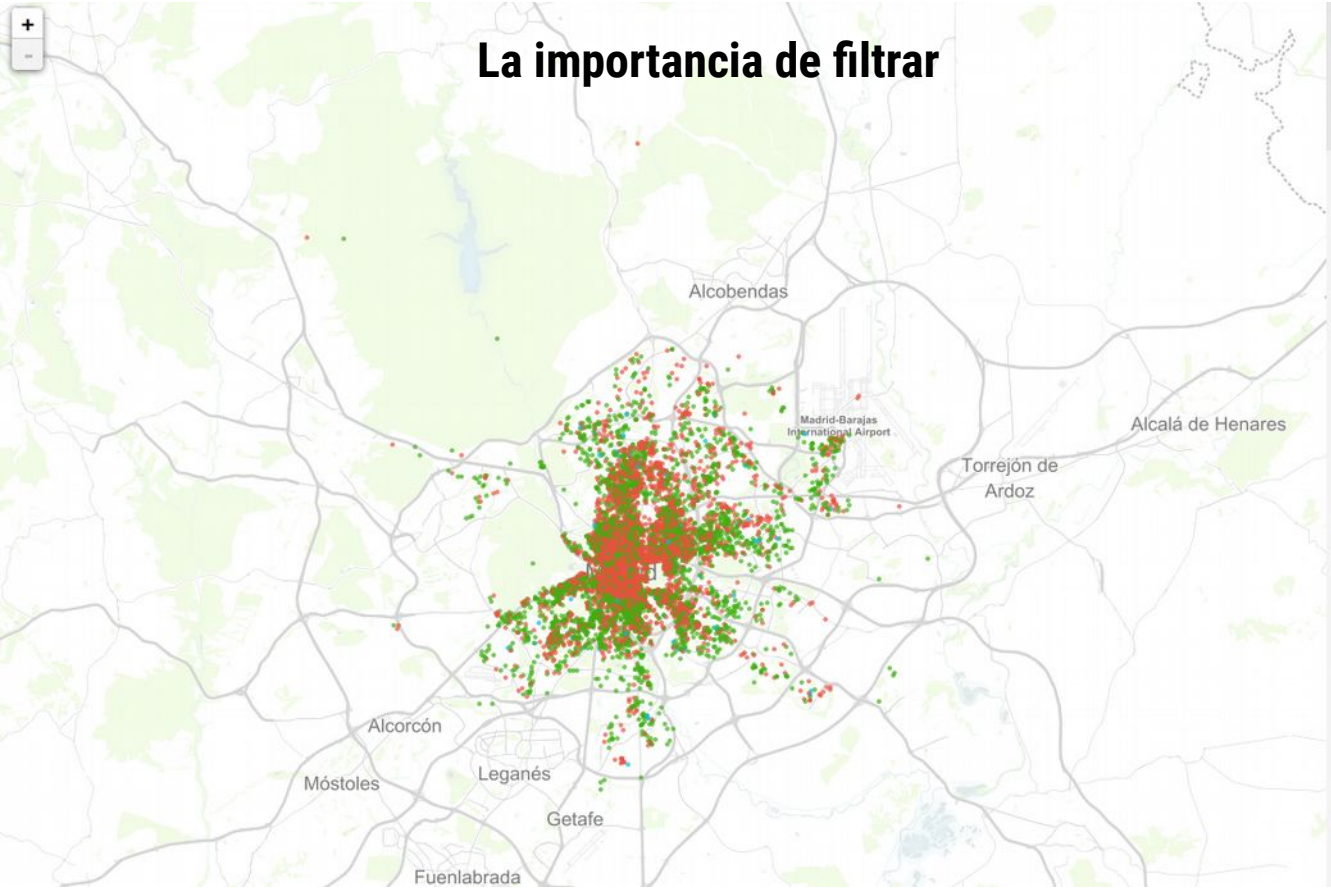
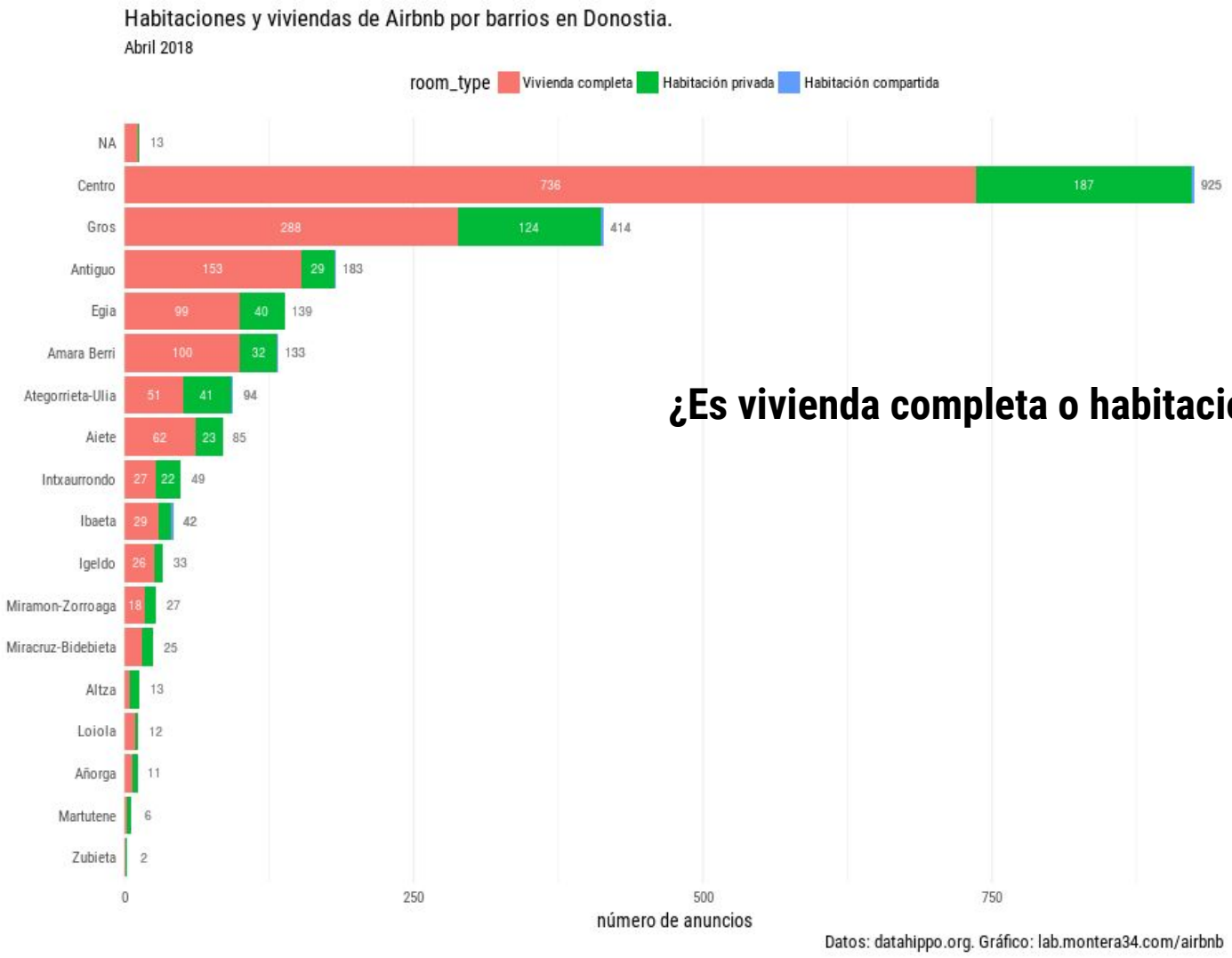


Imagen: http://insideairbnb.com/madrid/



¿Es vivienda completa o habitación?

¿Cuántos anuncios tienes? ¿Multipropietario? ¿multigestor?

Ayuda a distinguir el tipo de uso que se le da al alojamiento: esporádico vs comercial.

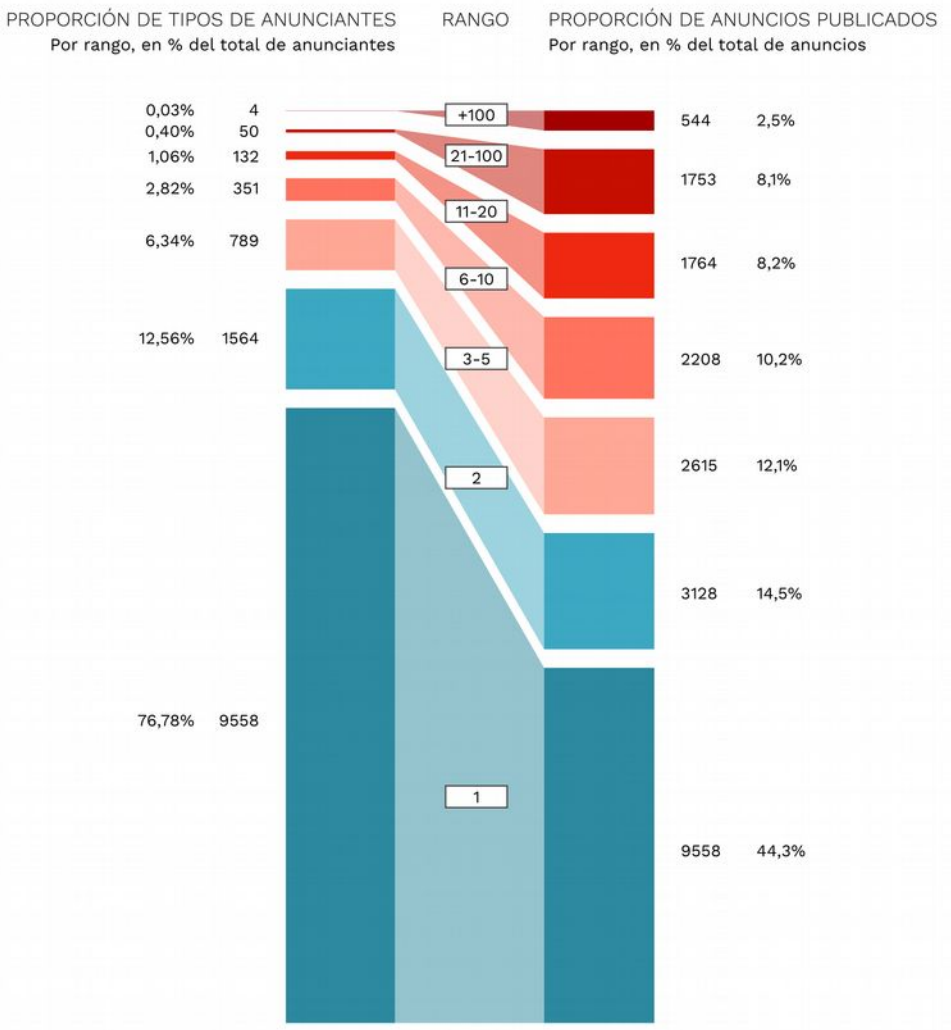


Imagen: Jorge Toledo, Ibai Zabaleta y Diego Casado
https://somosmalasana.elperiodico.com/las-empresas-que-co

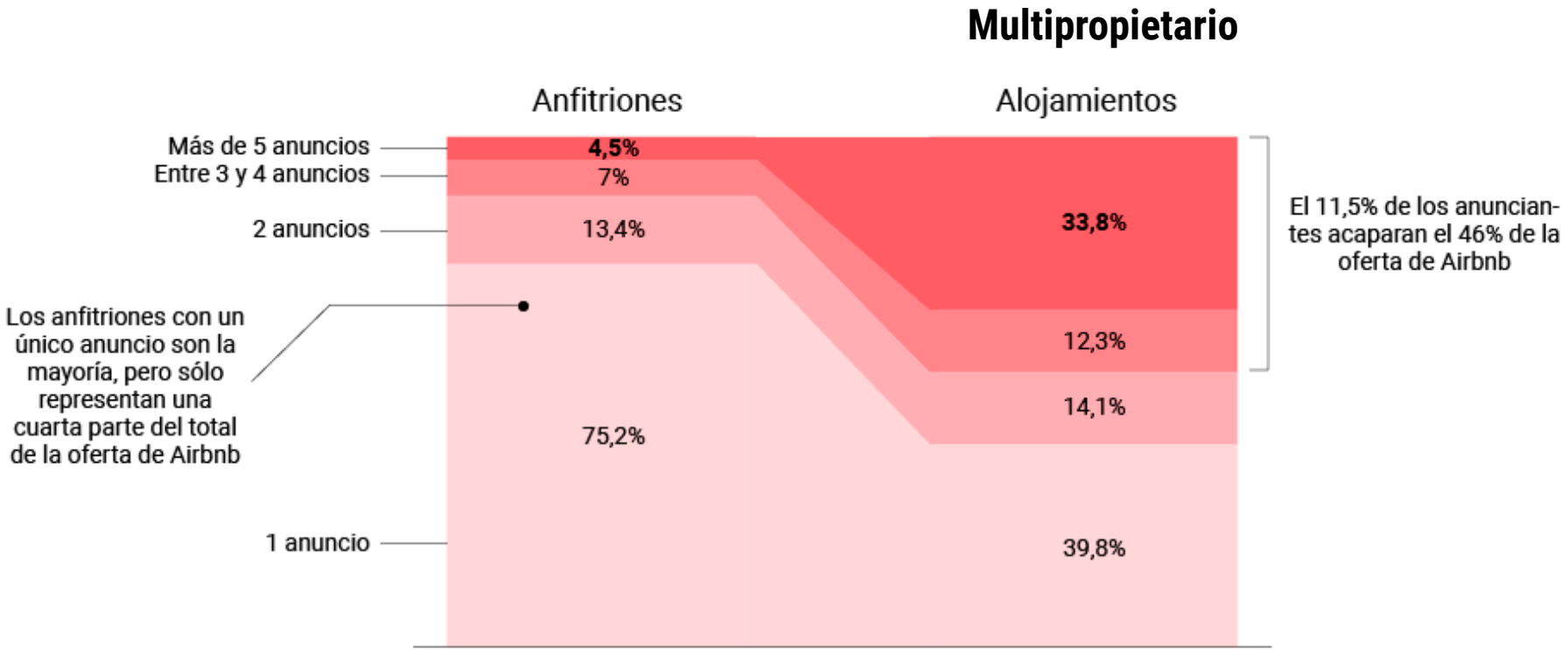
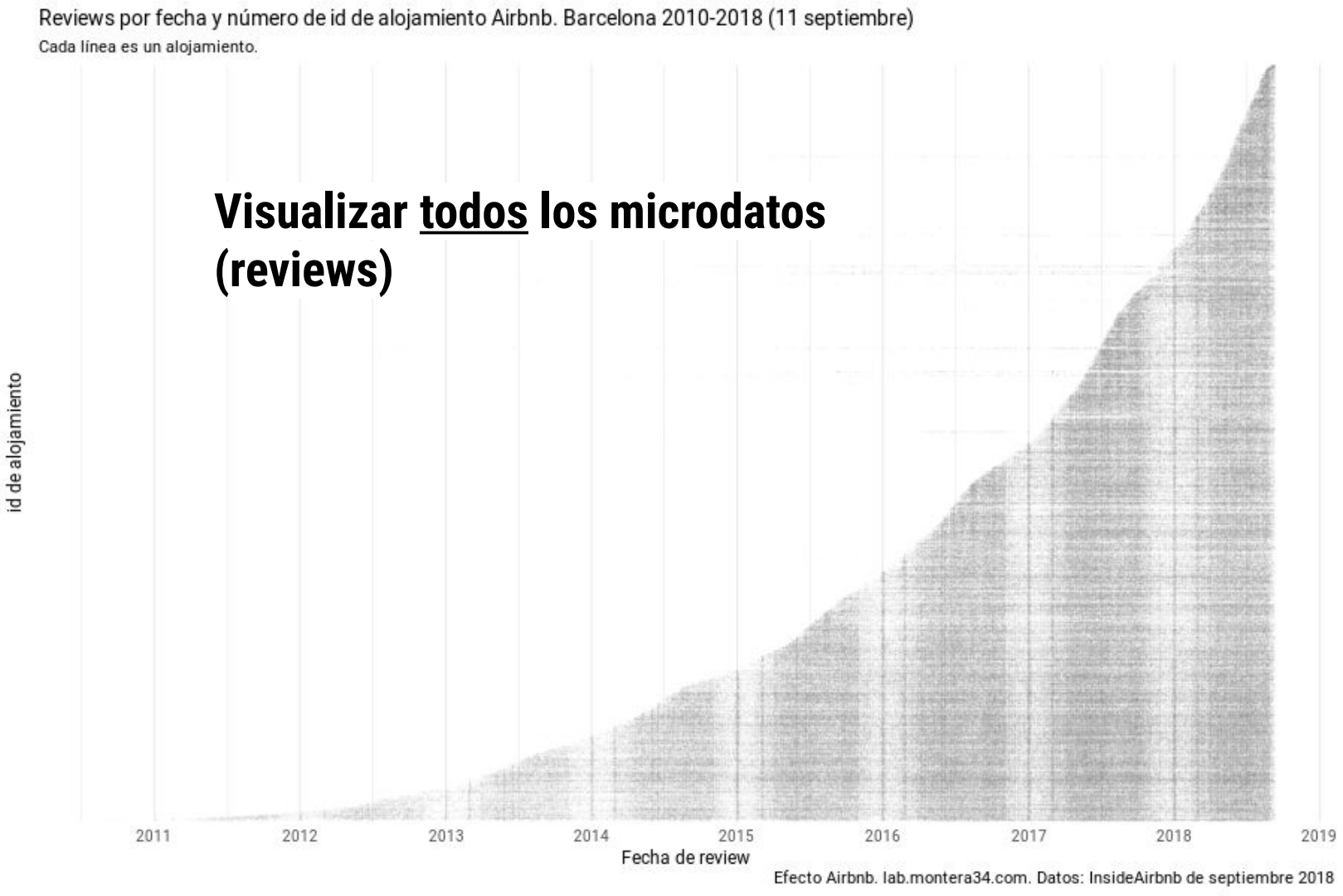
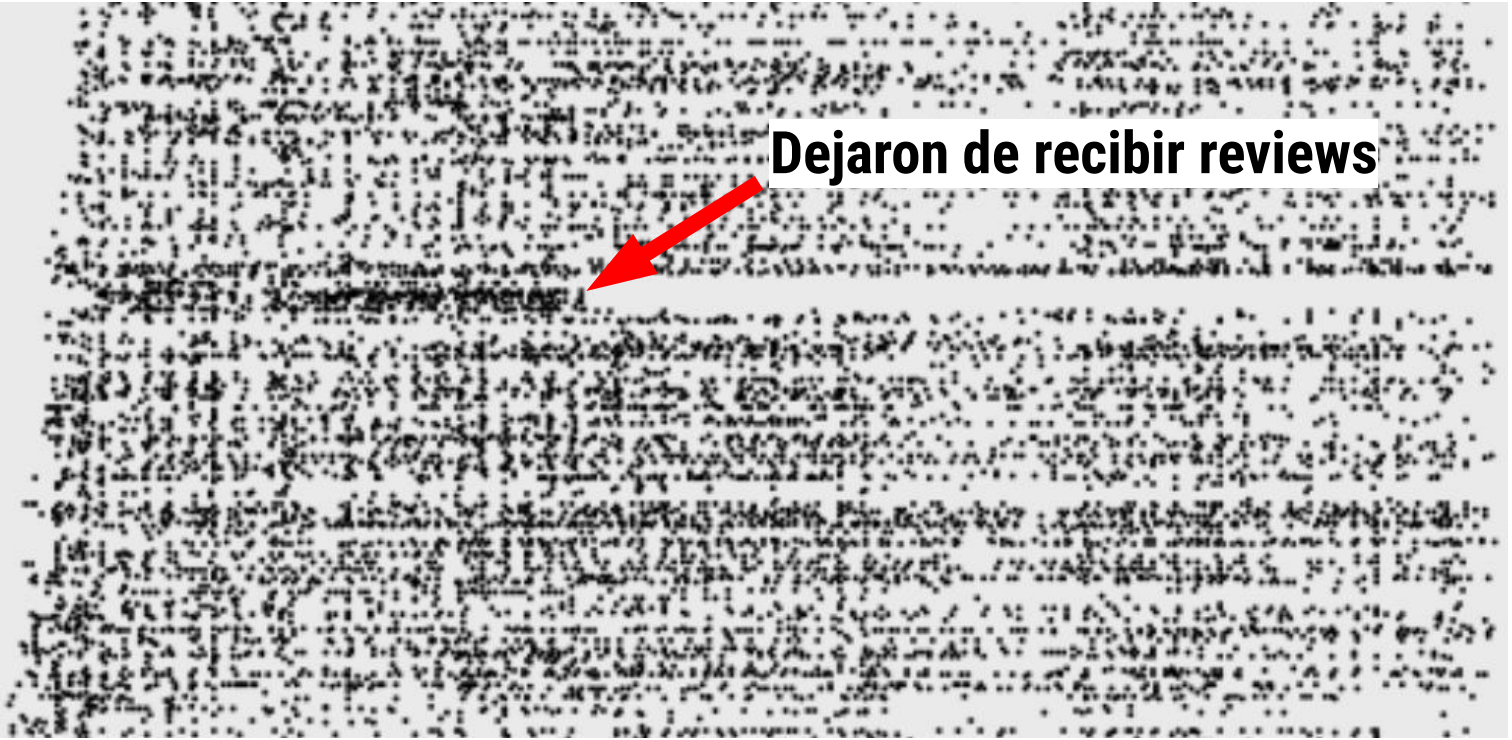


Imagen: Raúl Sánchez y Ana Ordaz
https://www.eldiario.es/economia/Airbnb-Espana-anfitriones-gestionan-alojamientos_0_806669478.html


Reviews por fecha y número de id de alojamiento Airbnb. Barcelona 2010-2018 (11 septiembre)
Cada línea es un alojamiento.
Visualizar todos los microdatos (reviews)
id de alojamiento
2011
2012
2013
2014
2015
2016
2017
2018
2019
Fecha de review
Efecto Airbnb. lab.montera34.com. Datos: InsideAirbnb de septiembre 2018

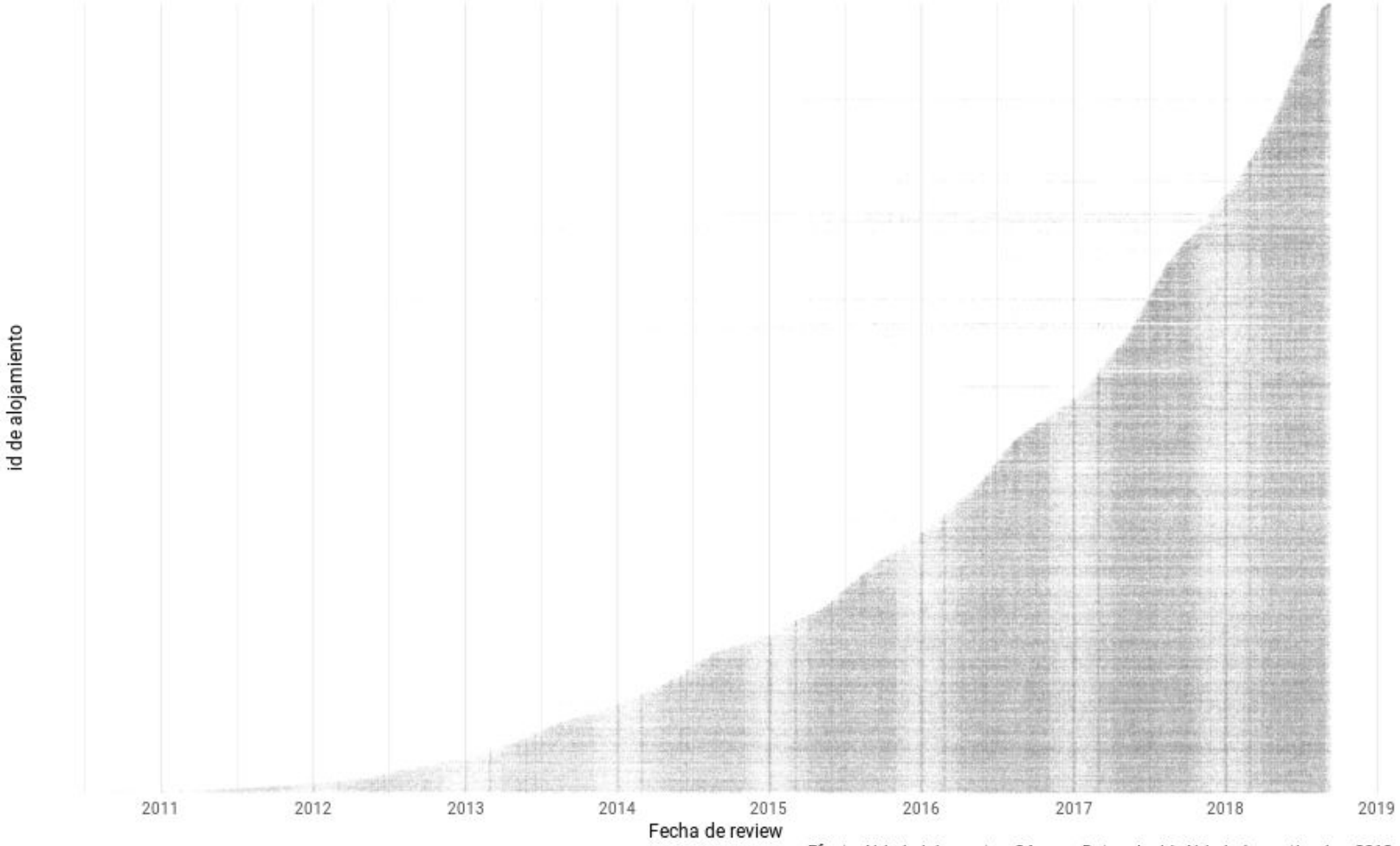
Cada línea es una anuncio, cada punto una review
El eje horizontal es el tiempo



Las visualizaciones plantean nuevas preguntas

Reviews por fecha y número de id de alojamiento Airbnb. Barcelona 2010-2018 (11 septiembre)

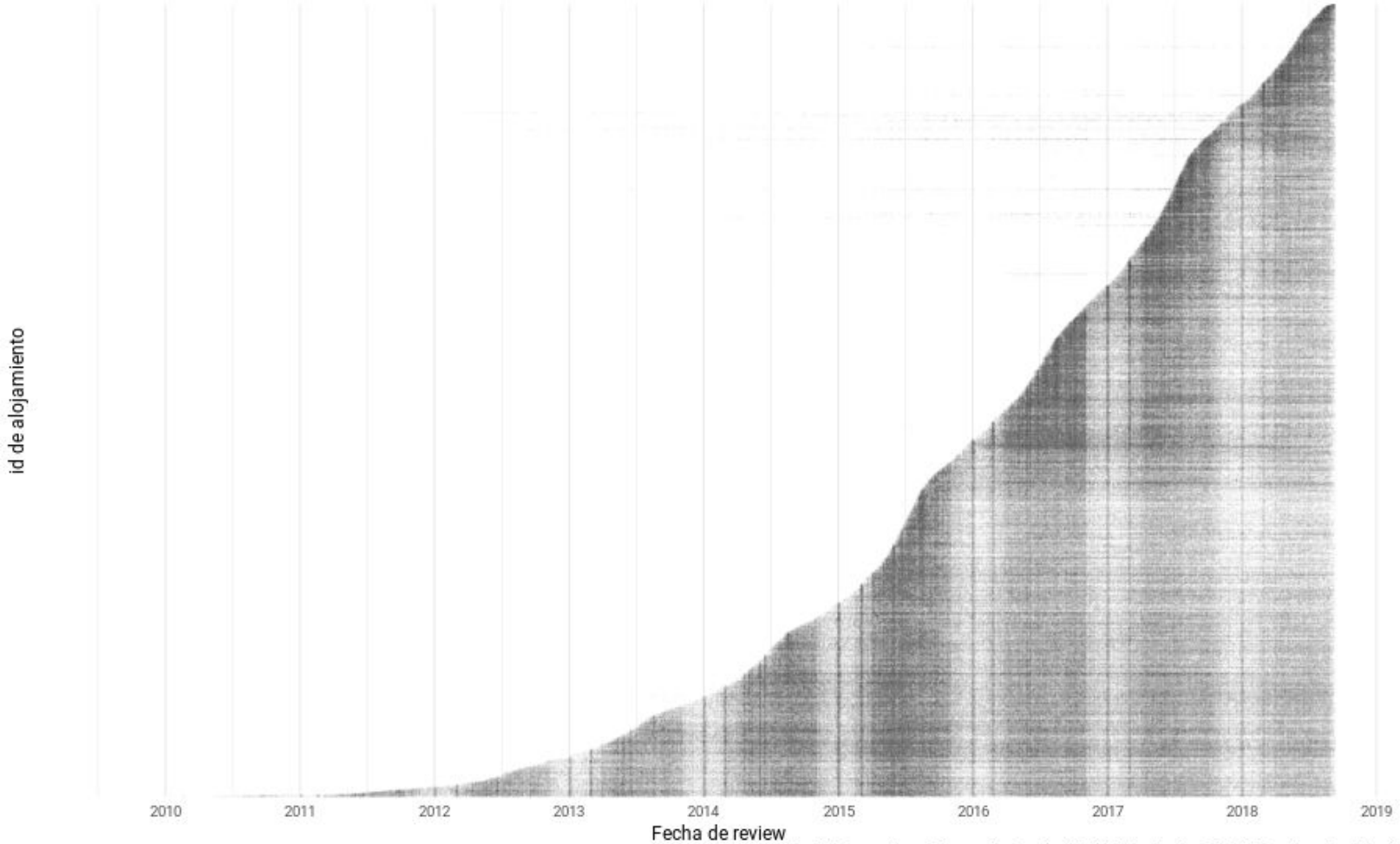
Cada línea es un alojamiento.



Efecto Airbnb. lab.montera34.com. Datos: InsideAirbnb de septiembre 2018

Reviews por fecha y número de id de alojamiento Airbnb. Barcelona 2010-2018 (11 septiembre 2018)

Cada línea es un alojamiento.



Efecto Airbnb. lab.montera34.com. Datos: InsideAirbnb desde 2015 (27 series de datos)

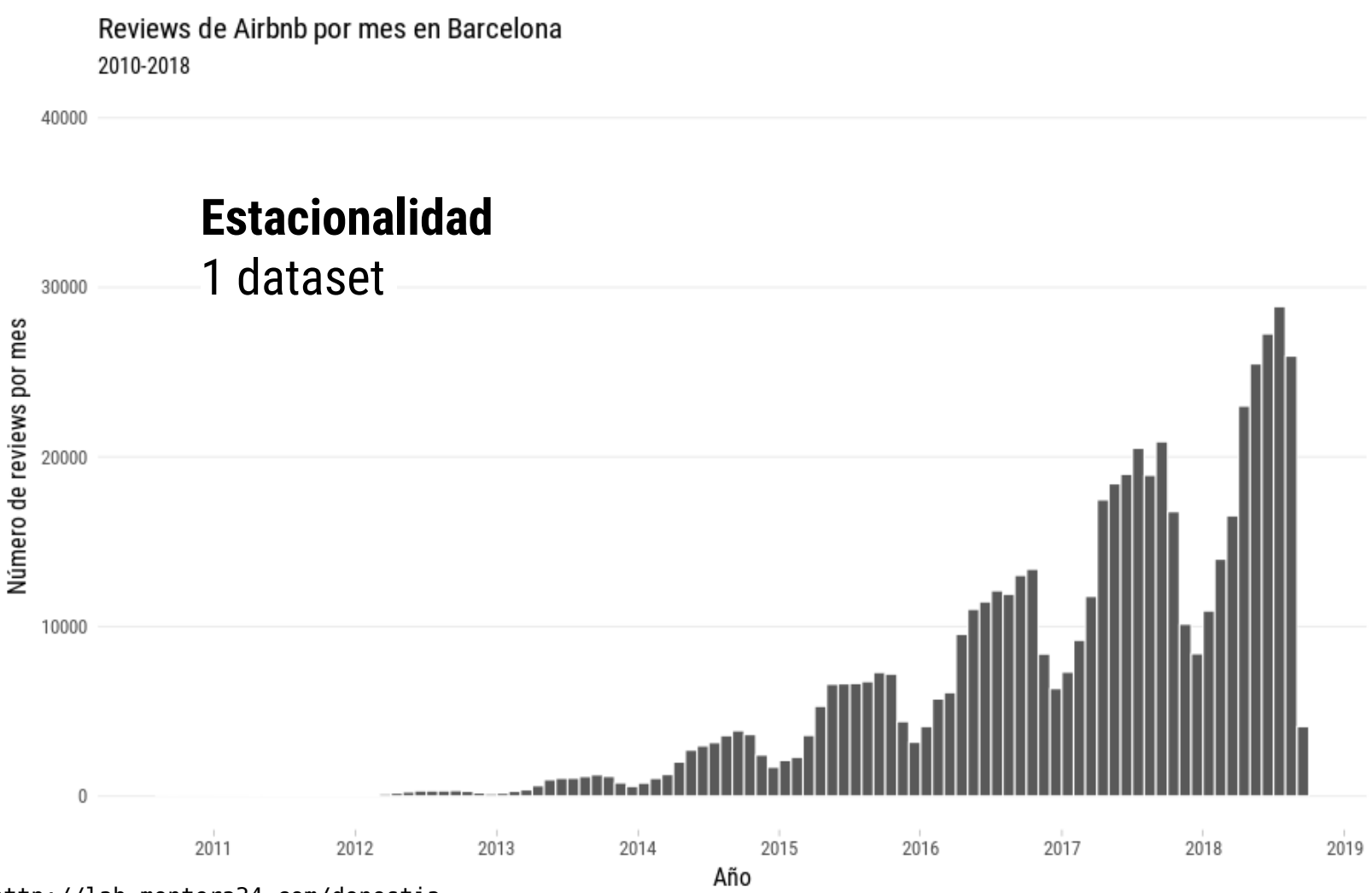


Imagen: http://lab.montera34.com/donostia

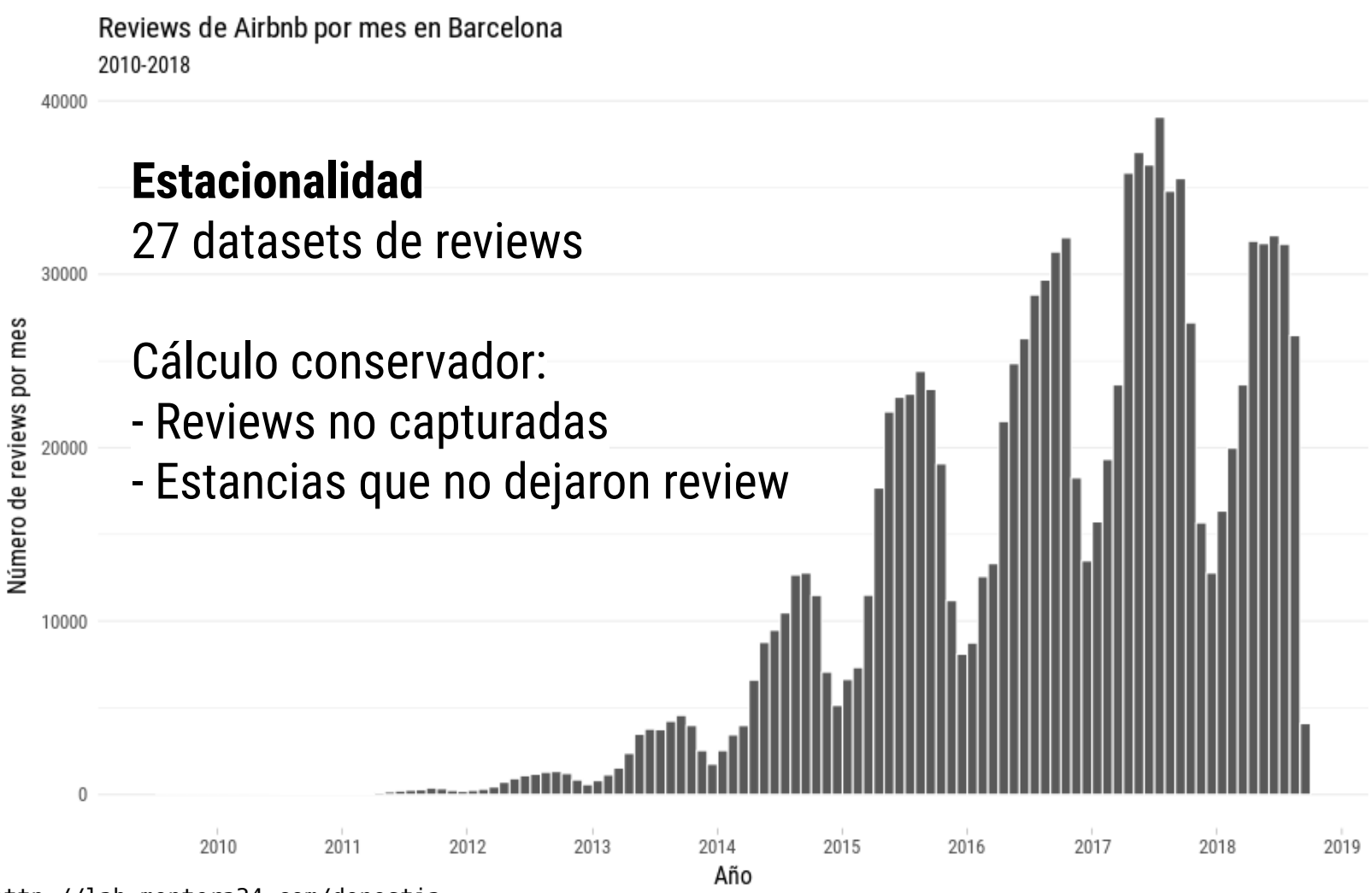
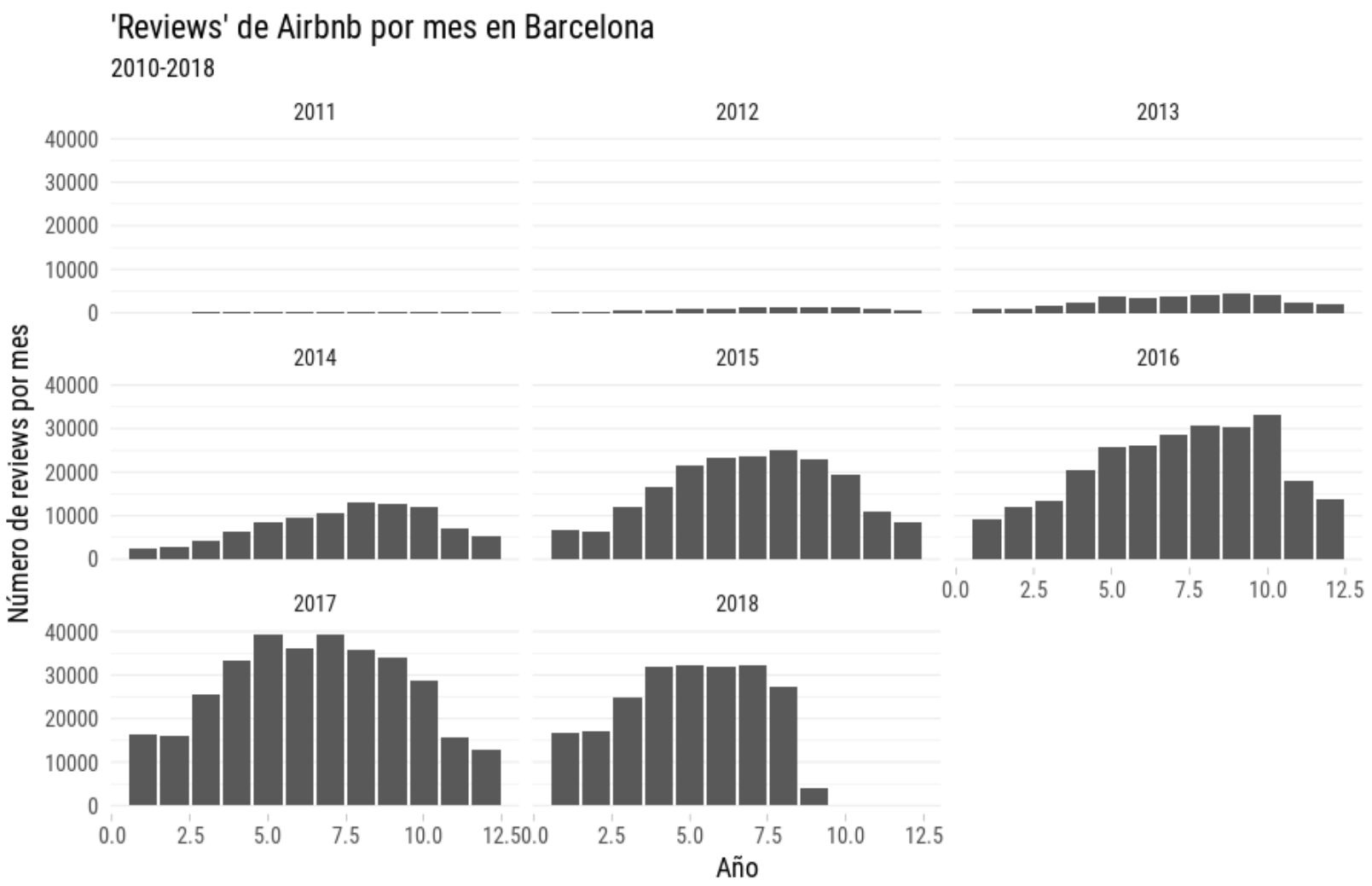
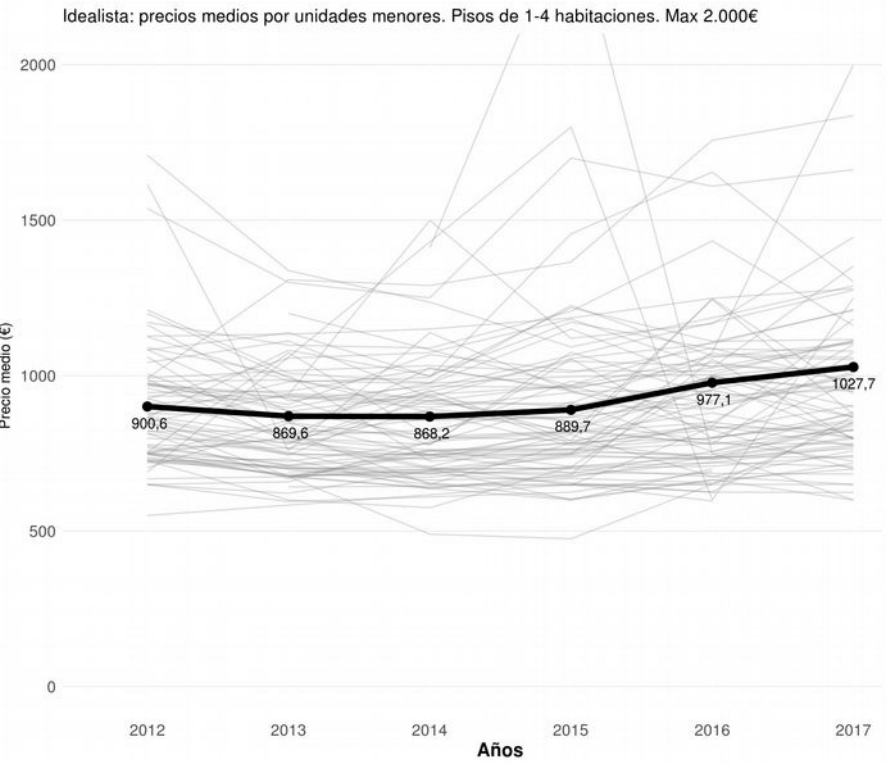


Imagen: http://lab.montera34.com/donostia


'Reviews' de Airbnb por mes en Barcelona
2010-2018
2011
2012
2013
2014
2015
2016
2017
2018
Número de reviews por mes
Año
Efecto Airbnb. lab.montera34.com. Datos: 26 archivos de InsideAirbnb desde 2015

El problema de no tener los microdatos: Las medias, las medianas y agregaciones



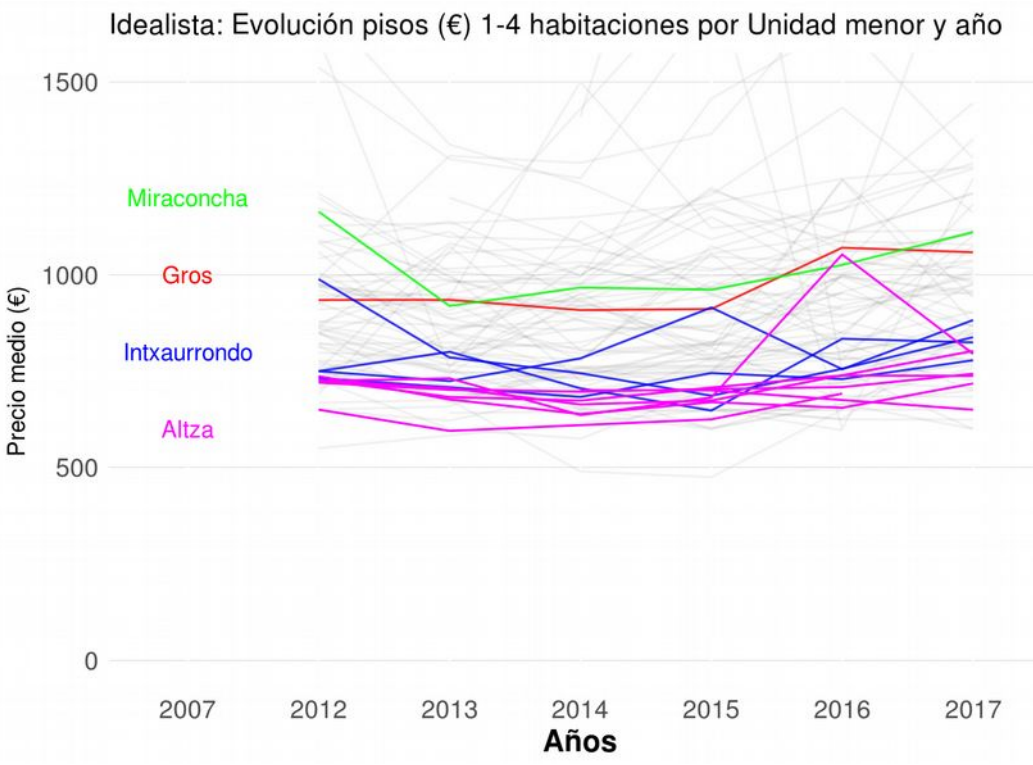


Imagen: Montera34 para Donostiako Etxeginza

Visualizando la incertidumbre

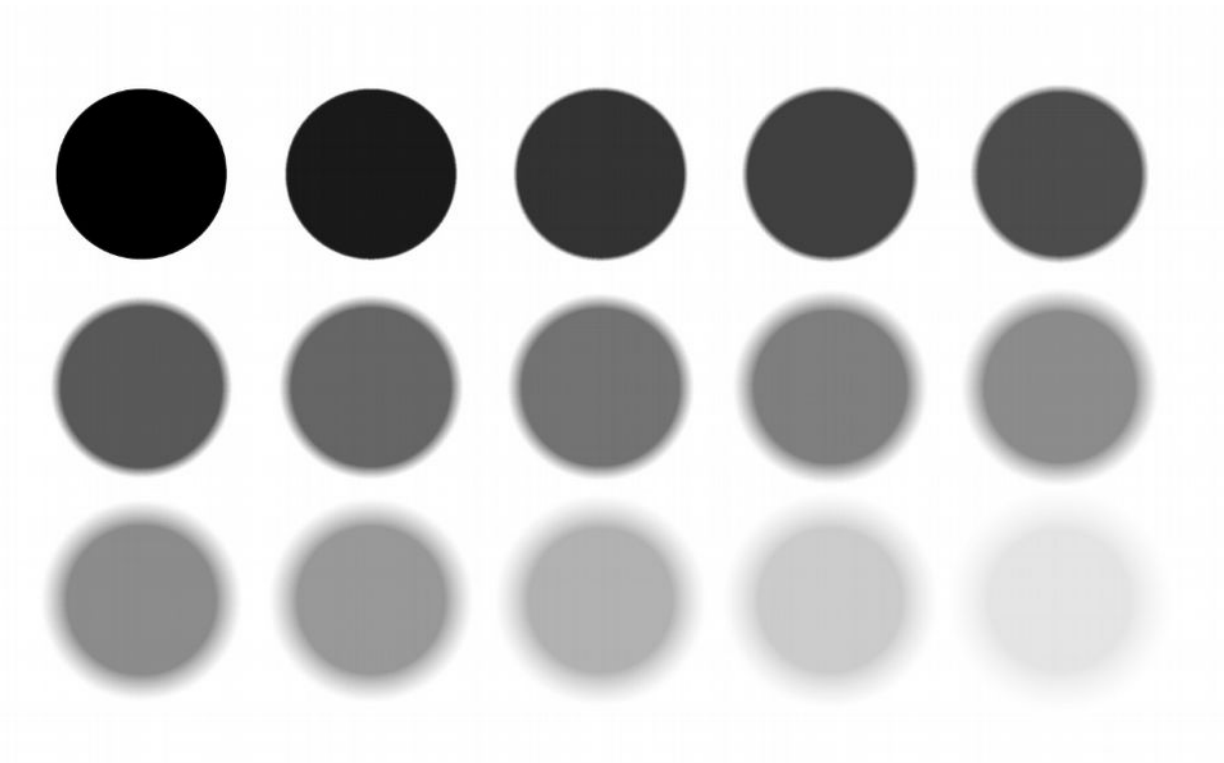


Imagen: https://flowingdata.com/2018/01/08/visualizing-the-uncertainty-in-data/

Visualizando la incertidumbre

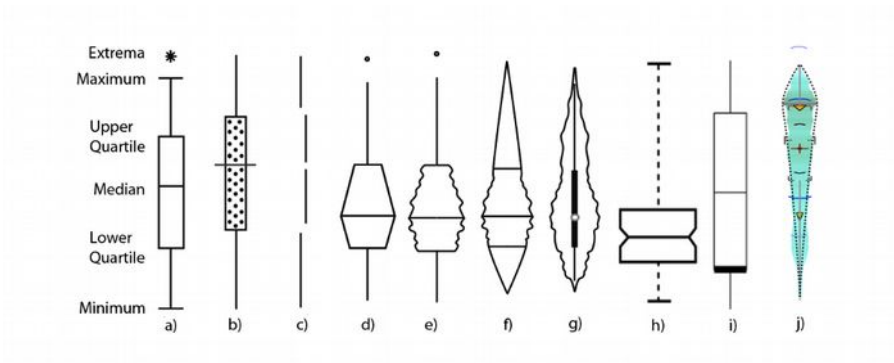


Fig. 1.2 Variations of the boxplot. a) The construction of the boxplot [106] b) Range plot [97] c) Innerquartile plot [105] d) Histplot [4] e) Vaseplot [4] f) Box-percentile plot [26] g) Violin plot [40] h) Variable width notched boxplot [68] i) Skewplot [16] j) Summary plot [85]

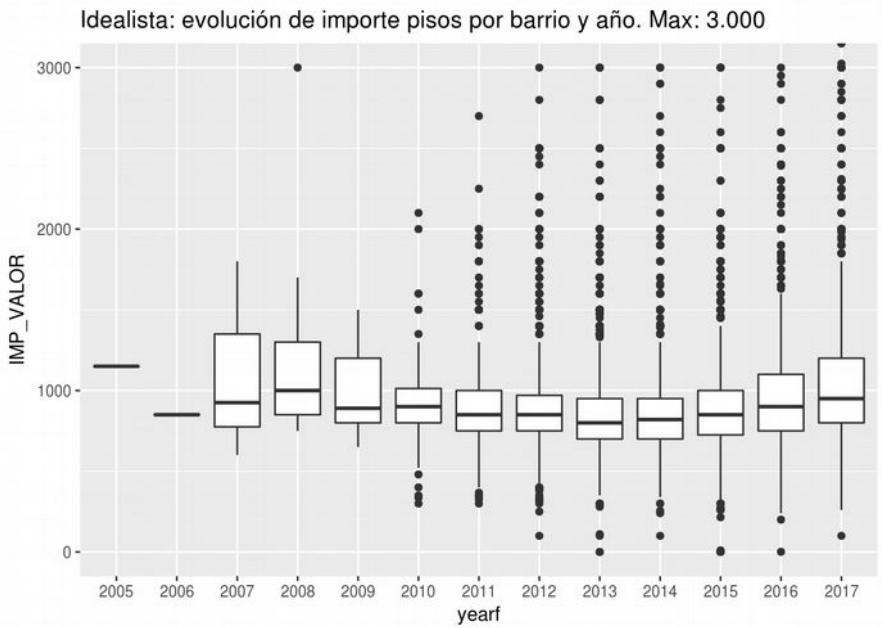


Imagen: http://www.sci.utah.edu/publications/Bon2014a/Overview-Uncertainty-Visualization-2015.pdf
Montera34 para Donostiako Etxeginza

Indicadores de turistificación

Indicadores para poder comparar regiones, ciudades...

Densidades de superficie:

- Anuncios / hectárea
- Pisos completos / hectárea
- Plazas / hectárea

Ratios:

- Plazas / plazas hoteleras
- Anuncios / viviendas
- Pisos completos / viviendas
- Plazas / habitantes

Imagen: https://lab.montera34.com/airbnb/barrios/

Ayúdanos a documentar

Preguntas, ideas y deseos

¿Cómo podemos colaborar?
¿Qué retos investigar?
¿Qué preguntas plantear?
¿Qué datos necesitamos?
  ¿Cuáles podemos conseguir y/o pedir?
¿Cómo visualizar?
¿Cómo explicar?

¡Por un debate informado!

#data

Toda la info y links en:
montera34.com/bcn2018

Simposio: viernes mañana.
- Intro
- Ronda presentaciones
- Ponencias
- Debate

Taller: viernes tarde, sábado mañana y domingo mañana.