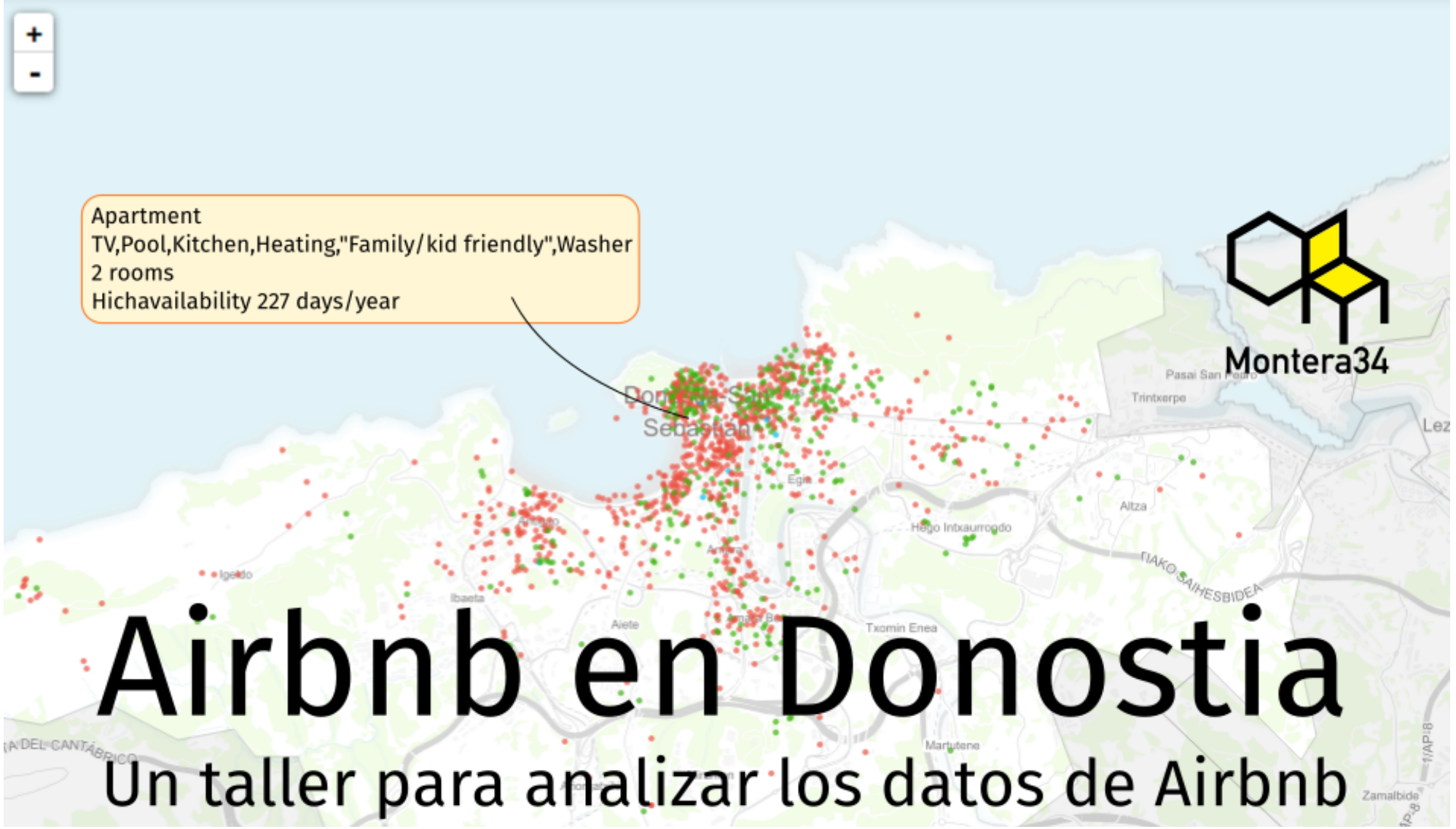


Airbnb en Donostia

Un taller para analizar los datos de Airbnb

El efecto de Airbnb en Donostia-San Sebastián
Taller para analizar los datos de Airbnb

3-5 abril 2017. Hirikilabs. Tabakalera. Donostia.

Web: https://montera34.com/project/efecto-airbnb-donostia/

#airbnbdonosti

montera34.com

Alfonso Sánchez Uzábal y Pablo Rey Mazón

Objetivos taller

Aprender a usar herramientas para producir un informe basado en datos geoespaciales.

Plantear preguntas y responderlas con análisis y visualización.

Aprender técnicas y herramientas.

Estructura del taller

Día 1: lunes 3 de abril

Presentaciones 30min
- Del taller. Qué vamos a hacer.
- De Montera34. Quiénes somos. Qué hacemos.
- De los talleristas. 2 tuits.

Sobre visualización de datos y mapas 1h30
- Algo de teoría. Pasos para generar visualización
- Algunas herramientas: Openrefine, Pivot tables en Google Drive, Carto.com.

Fairbnb. Por Sito Veracruz. 1h

Día 2-3. Martes 4-5 de abril

Taller de análisis y visualización colectivo de datos.
Encontrar respuesta a las preguntas planteadas.

Quiénes somos

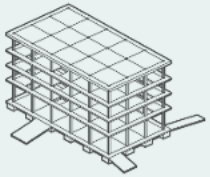
Montera34 somos
Alfonso S. Uzábal (@skotperez, voragine.net). Arguilla
Pablo Rey Mazón (@numeroteca, numeroteca.org). Bilbao

info@montera34.com
@montera34
Montera34.com

cadaveresinmobiliarios.org/mapa/

enlace_web, enlaces_noticias, material_grafico, archi

Mapa de cadáveres inmobiliarios



Mapa de proyectos arquitectónicos y desarrollos urbanísticos inacabados, infrautilizados o vacíos del periodo post-burbuja inmobiliaria española.

Este mapa es una de las visualizaciones creadas a partir de la base de datos de cadáveres inmobiliarios.

Sobre el proyecto Cadáveres Inmobiliarios

Ante la inexistencia de un inventario oficial, **Cadáveres Inmobiliarios: Base de datos post-burbuja** nace como un proyecto colectivo que persigue la localización y documentación de todos estos desarrollos que murieron antes de tiempo. Esta iniciativa es de marcado carácter interdisciplinar, englobando arquitectos, ingenieros, urbanistas, investigadores, artistas o activistas medioambientales e invita a la participación al mayor número de personas posible.

El objetivo de esta ambiciosa recopilación es la de proporcionar información exhaustiva de manera accesible para que cada cual la utilice acorde con sus intereses. Aunque el mero hecho de informar sobre este fenómeno supone una crítica implícita al modelo de desarrollo basado en el ladrillo, 'Cadáveres Inmobiliarios' pretende ser el punto de partida para futuros proyectos que propongan soluciones a situaciones hoy estancadas, fomentando así una concienciación social con respecto a una realidad que nos define y a la que no podemos dar la espalda.

search for address ...

Parque Empresarial 'Pedro Muñoz Industrial'

Leaflet | Map data © OpenStreetMap contributors, CC-BY-SA, Imagery © Mapb

BlacktodeFuture
@BlacktodeFuture

Siguiendo

Hace 10 años Antonio Cámara Eguinoa gastaba 121€ con su tarjeta black en DRINK-HANDICAP

10:00 - 3 abr. 2017

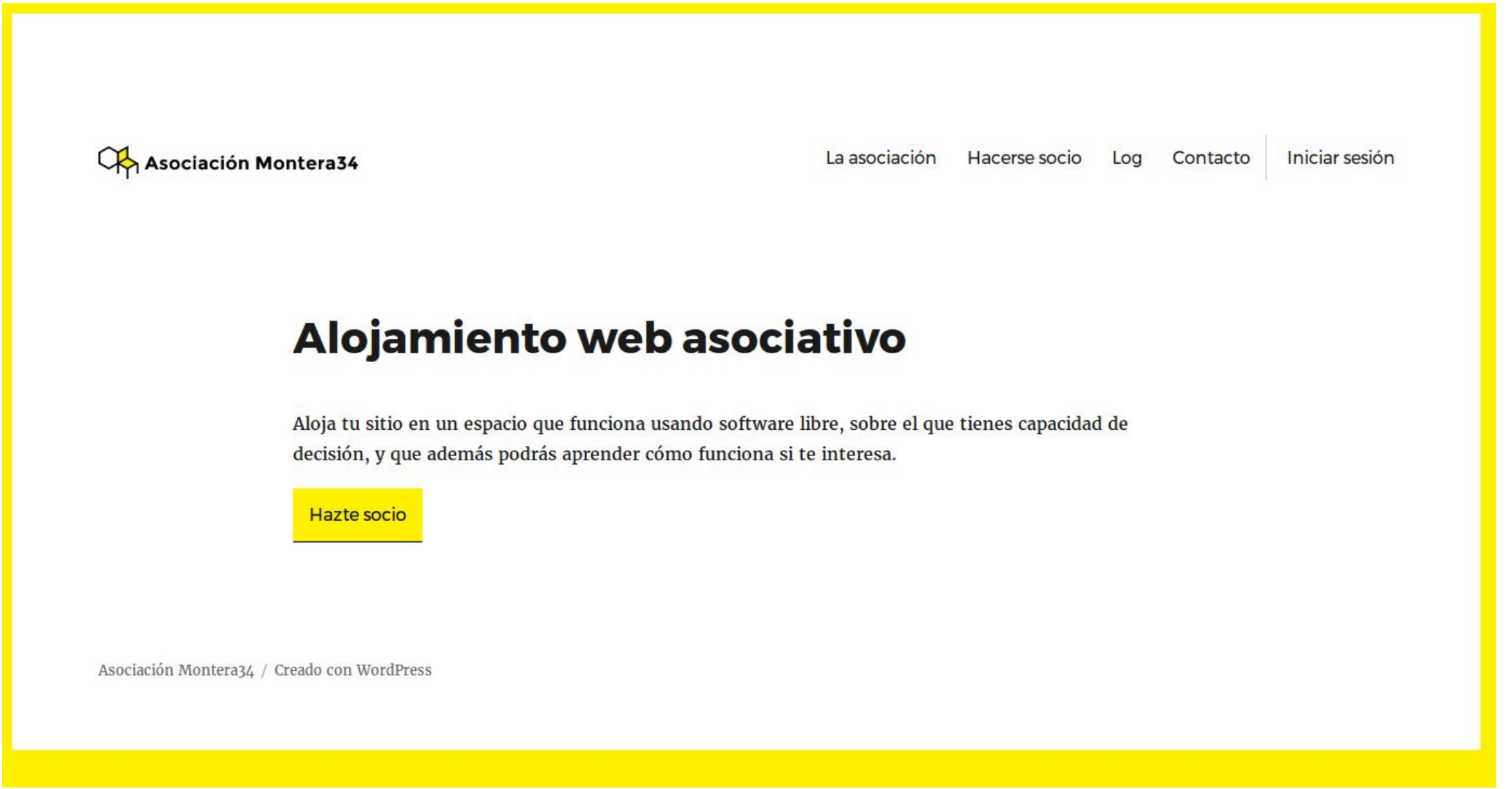
gastaba 34.73€ con su tarjeta black en ALIMENTACION ARMANDO

BlacktodeFuture @BlacktodeFuture · 4 h

Hace 10 años Javier de Miguel Sánchez gastaba 77.88€ con su tarjeta black en ROBLEGAS, S.A.

BlacktodeFuture @BlacktodeFuture · 5 h

Hace 10 años Rubén Cruz Uribe gastaba 74.59€ con su tarjeta black en FERMIN RUIZ ESCRIBANO


Asociación Montera34
La asociación
Hacerse socio
Log
Contacto
Iniciar sesión
Alojamiento web asociativo
Aloja tu sitio en un espacio que funciona usando software libre, sobre el que tienes capacidad de decisión, y que además podrás aprender cómo funciona si te interesa.
Hazte socio
Asociación Montera34 / Creado con WordPress

Captura de datos y análisis de red en Twitter

Mariluz Congosto

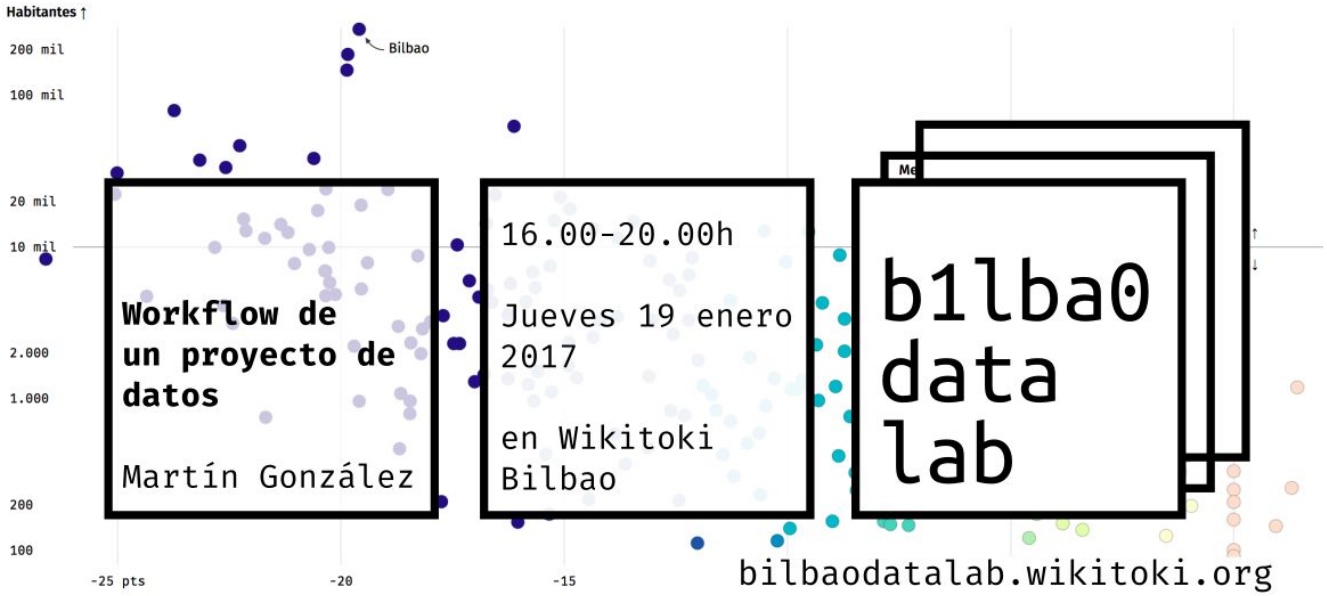
16.00-20.00h

Jueves 9 febrero 2017

en Wikitoki Bilbao

b1lba0 data lab

bilbaodatalab.wikitoki.org



Workflow de un proyecto de datos

Martín González

16.00-20.00h

Jueves 19 enero 2017

en Wikitoki Bilbao

b1lba0 data lab

bilbaodatalab.wikitoki.org

Presentación talleristas

Preséntate en dos tuits:

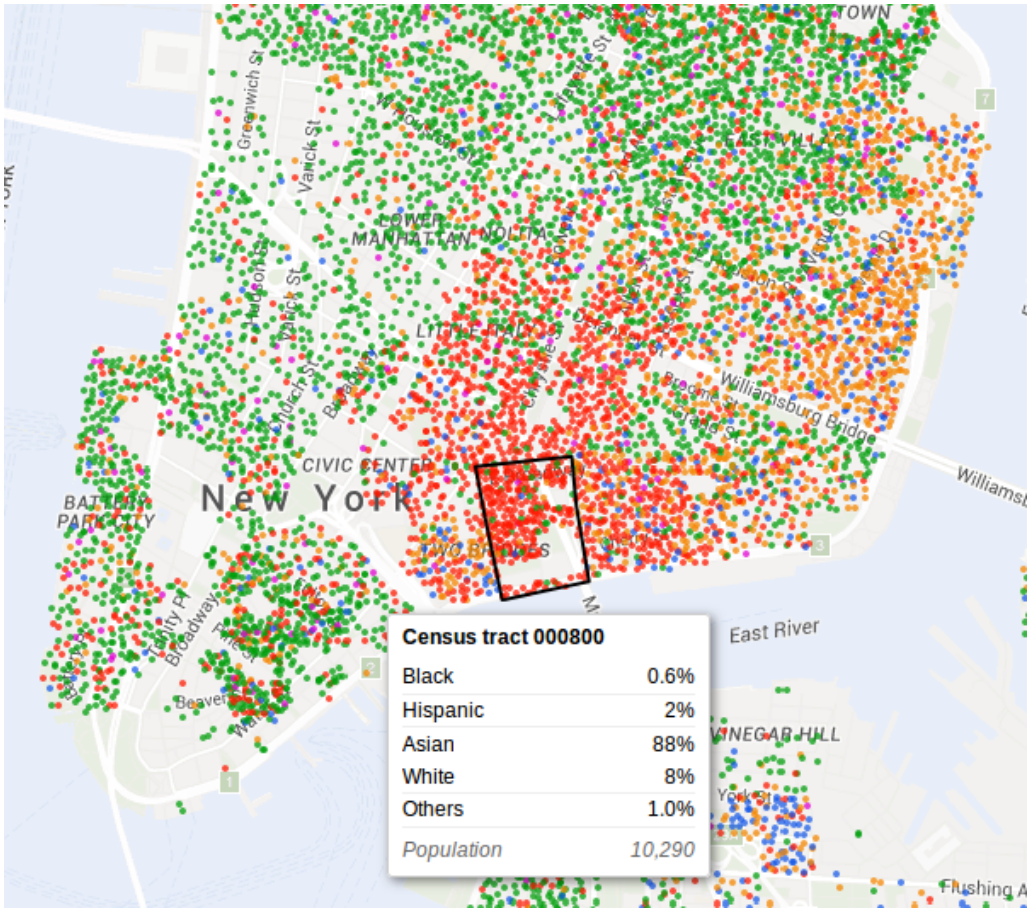
¿Quién eres?

¿Qué te ha motivado a apuntarte al taller? ¿Alguna idea o visualización en concreto que te gustaría desarrollar durante el taller?

Visualizar datos

¿Por qué visualizar?
Capacidad de síntesis.
Relacionar varias variables.

¿Por qué no visualizar?
No siempre es necesario:
con una tabla nos vale.



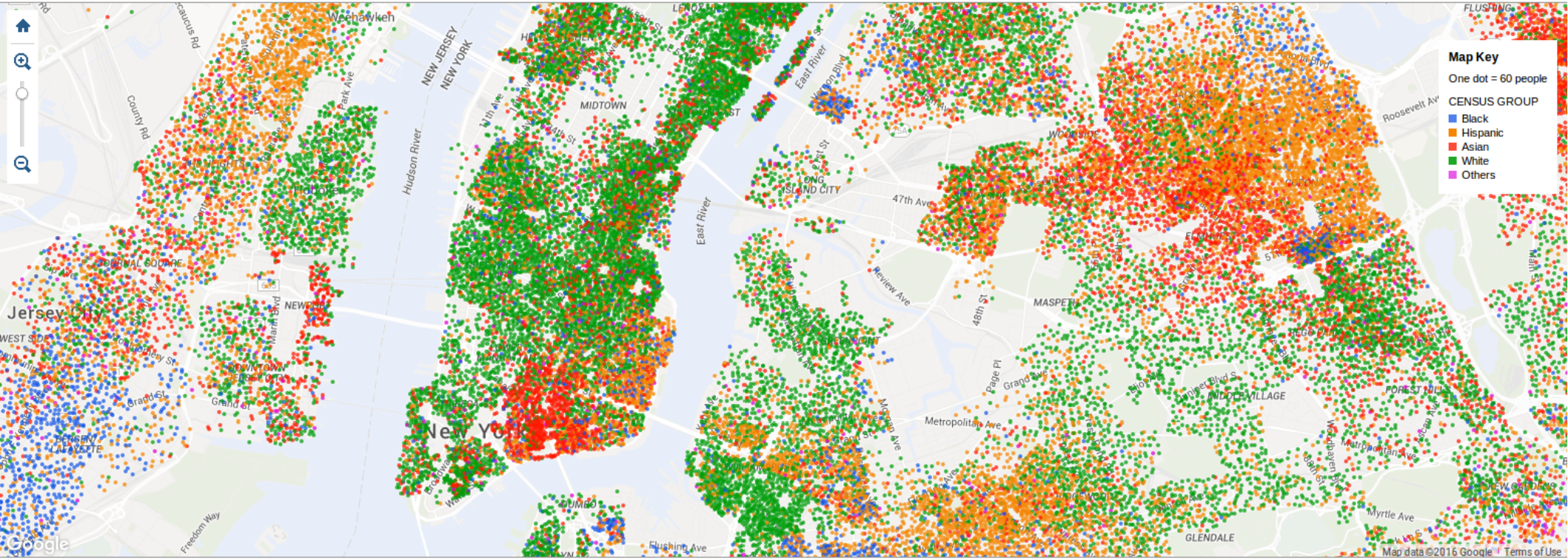
http://www.nytimes.com/interactive/2015/07/08/us/census-race-map.html

1 dot = 60 people

The New York Times
Published: July 8, 2015

Mapping Segregation

New government rules will require all cities and towns receiving federal housing funds to assess patterns of segregation.



Correction: In an earlier version of this map, the popup table misstated the share of each census group living in a census tract. For example, Manhattan tract 15200 is 56% white, not 54%.

By MATTHEW BLOCH, AMANDA COX and TOM GIRATIKANON

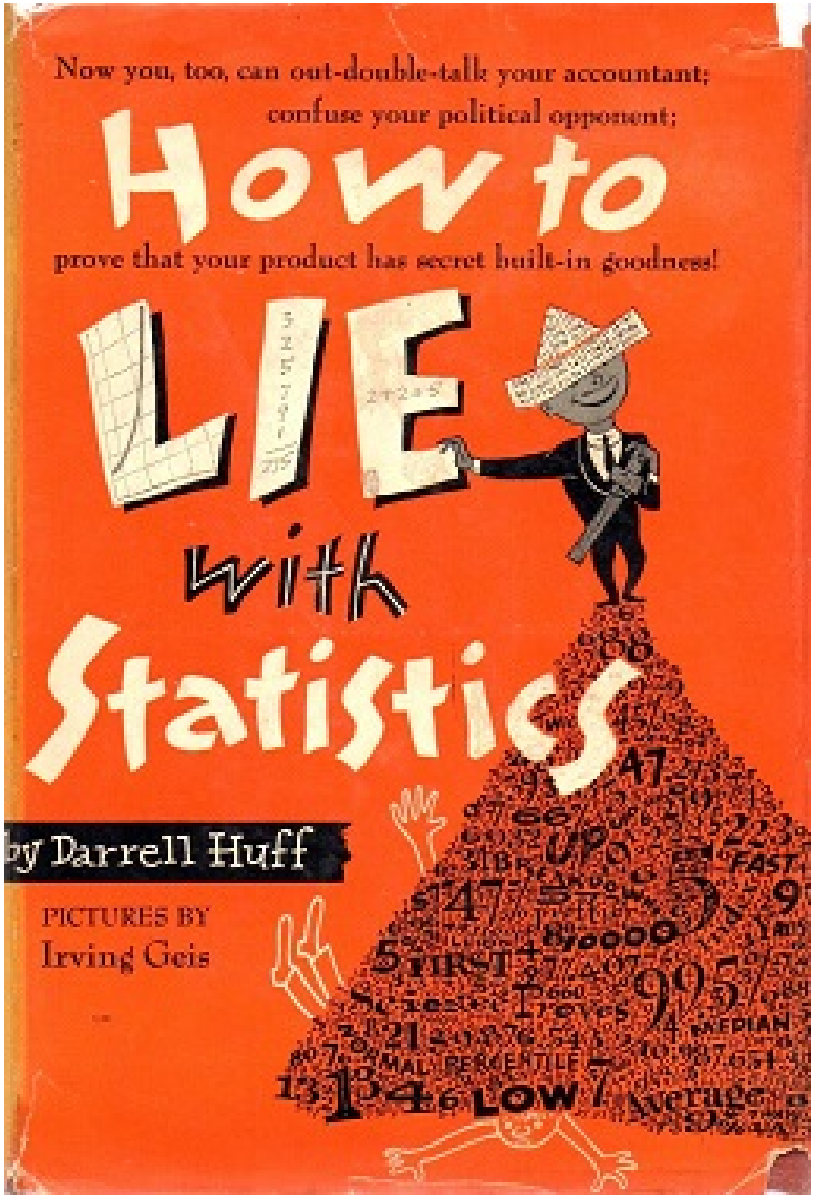
Sources: 2010 U.S. Census, socialexplorer.com, Google Maps

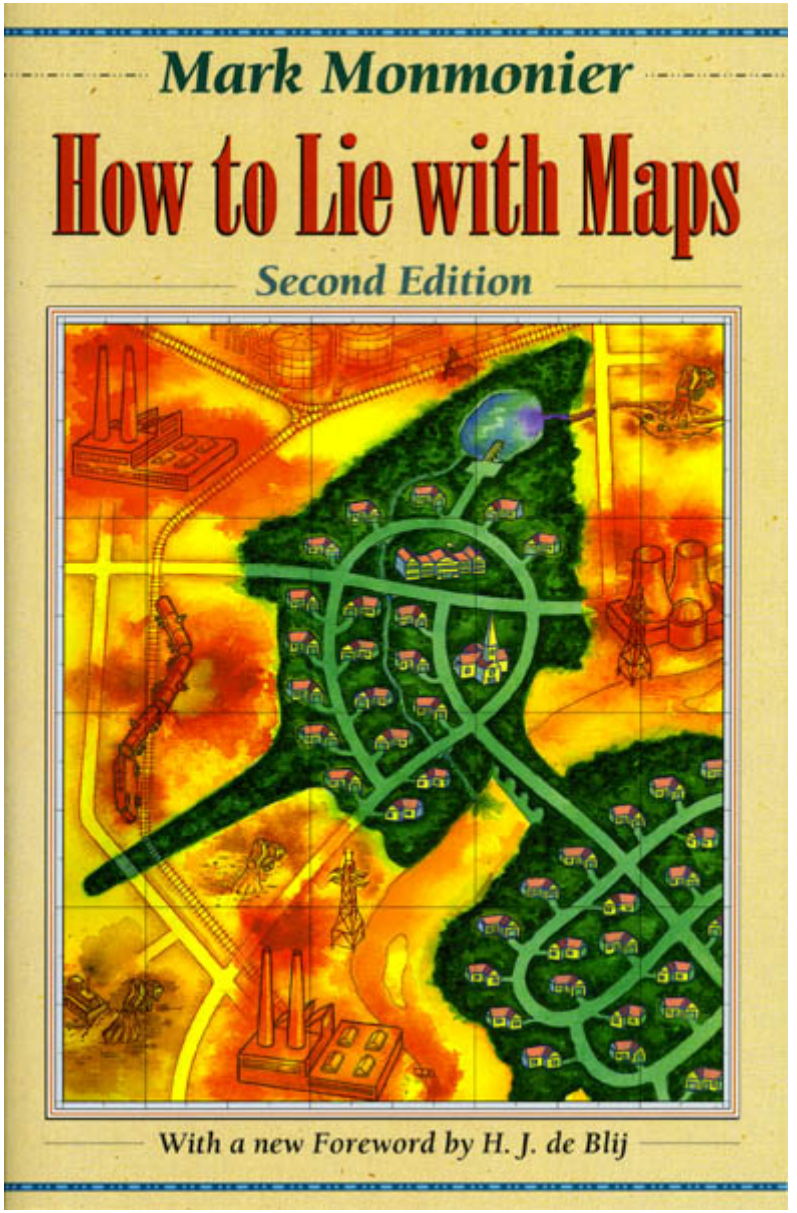
Subjetividad de las visualizaciones de datos

Los mapas no son neutrales ni objetivos.

La tecnología tampoco

http://numeroteca.org/2012/06/12/europa-2013/

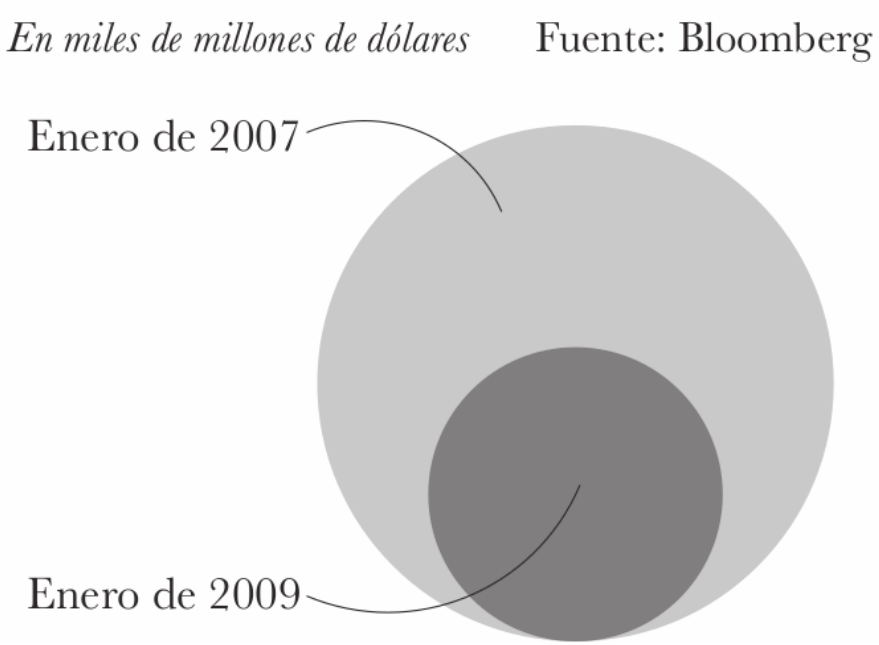




¿Cuánto mentimos?

Si no “mentimos“ o dejamos de mostrar información los mapas serían imposibles de comprender.

Societé Générale Market value



Example by Alberto Cairo. Visualizar. 2011

Tiempo medio de espera para una operación

A 30 de junio de 2016

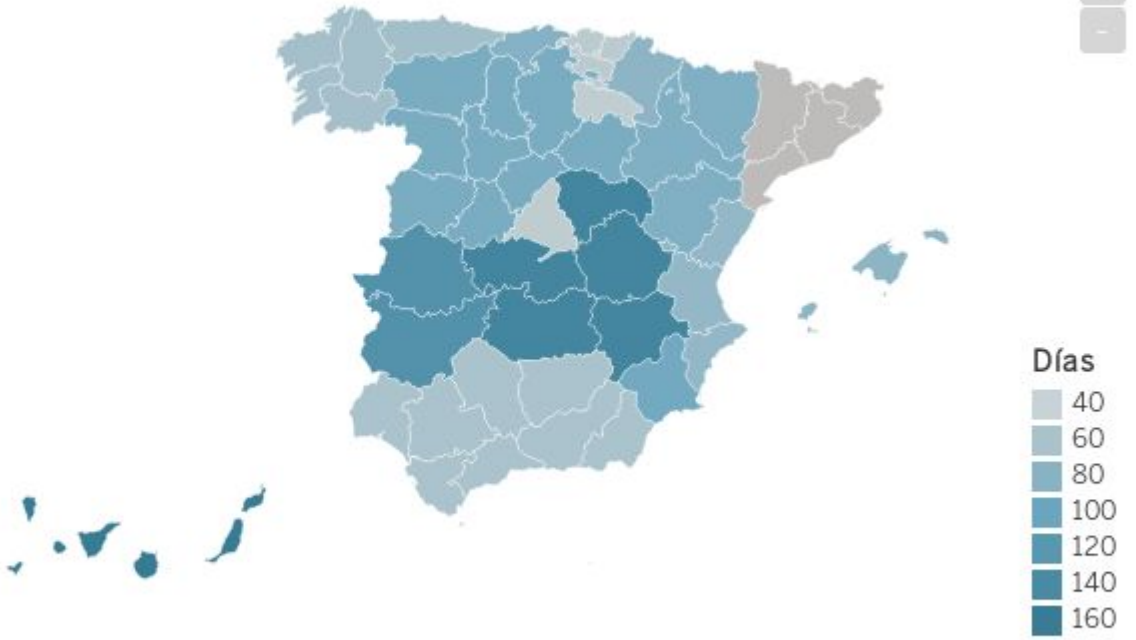
** No hay datos disponibles para Cataluña*

EL PAÍS / Fuente: Ministerio de Sanidad, Servicios Sociales e Igualdad

http://politica.elpais.com/politica/2017/03/30/actualidad/1490842694_426187.html

Tiempo medio de espera para una operación

A 30 de junio de 2016



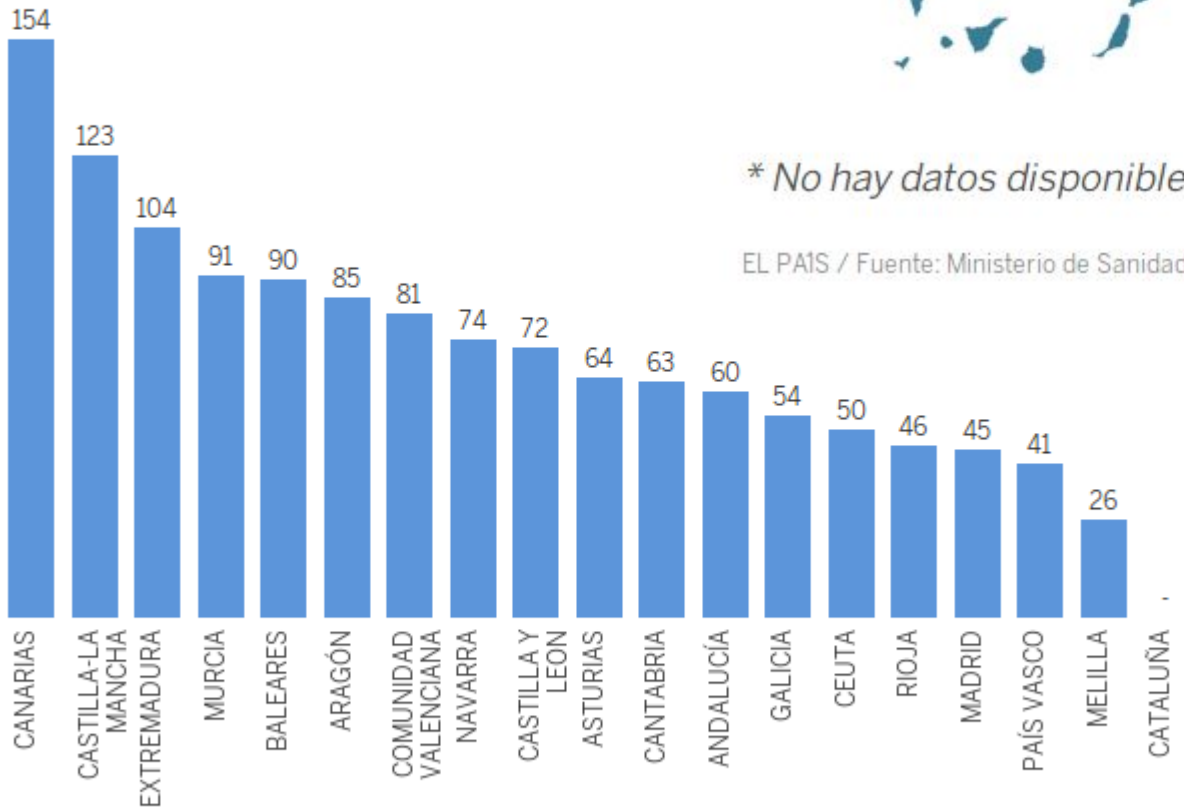
** No hay datos disponibles para Cataluña*

EL PAÍS / Fuente: Ministerio de Sanidad, Servicios Sociales e Igualdad

Espera quirúrgica por espec

En días, de media

Cirugía general y de digestivo



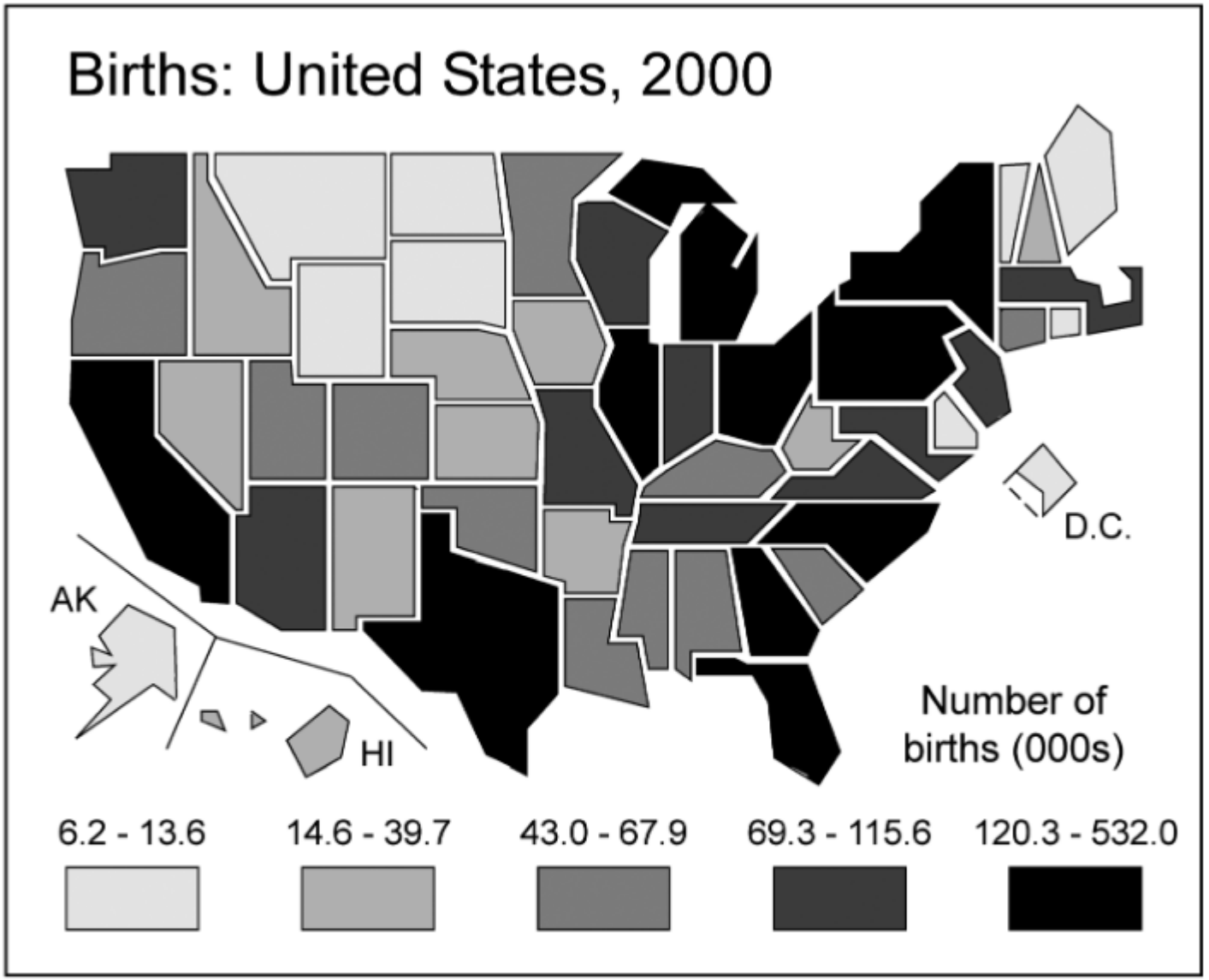


FIG. 7. *The darker-is-more-intense metaphor of choropleth maps offers a potentially misleading view of numbers of births.*

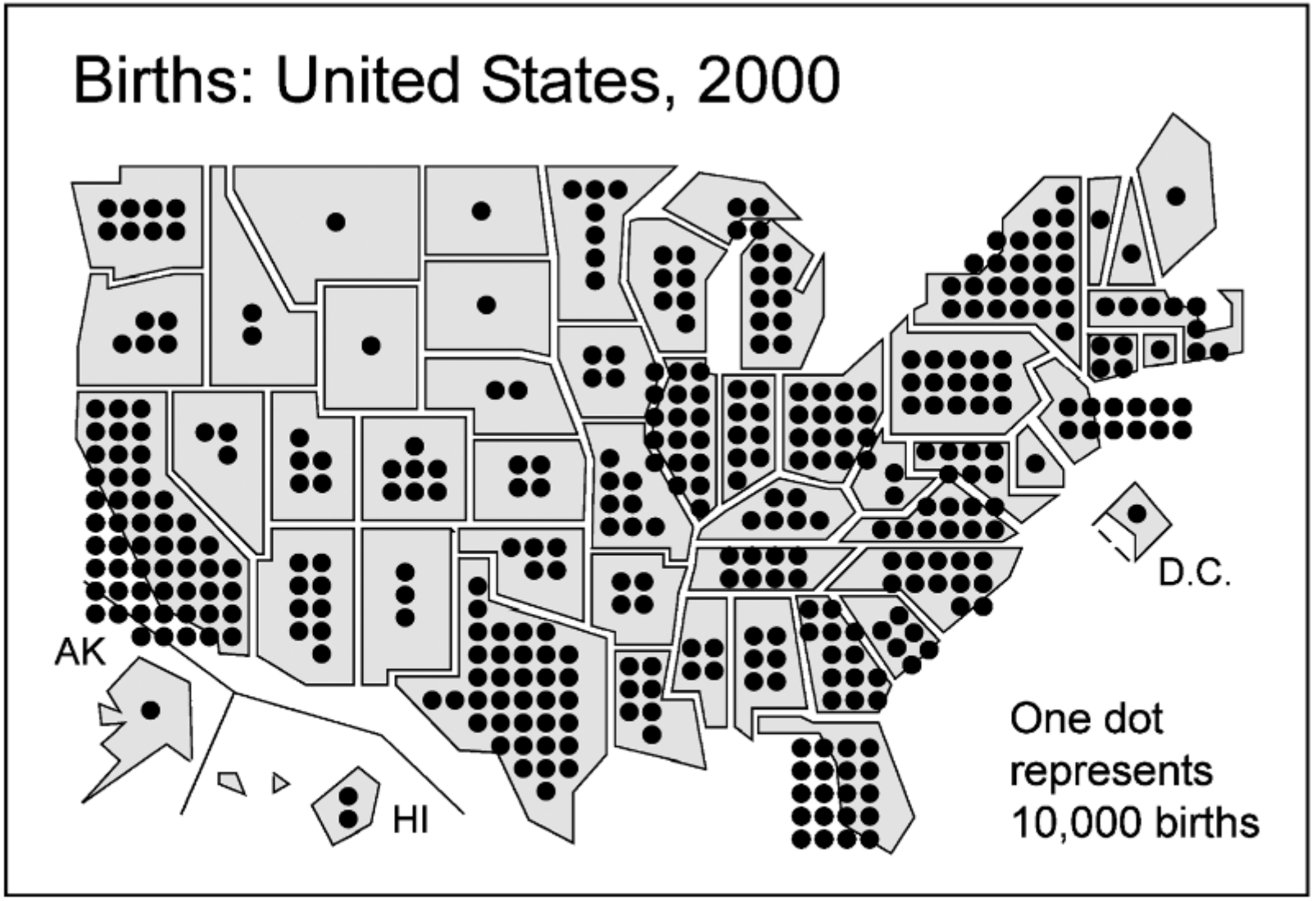
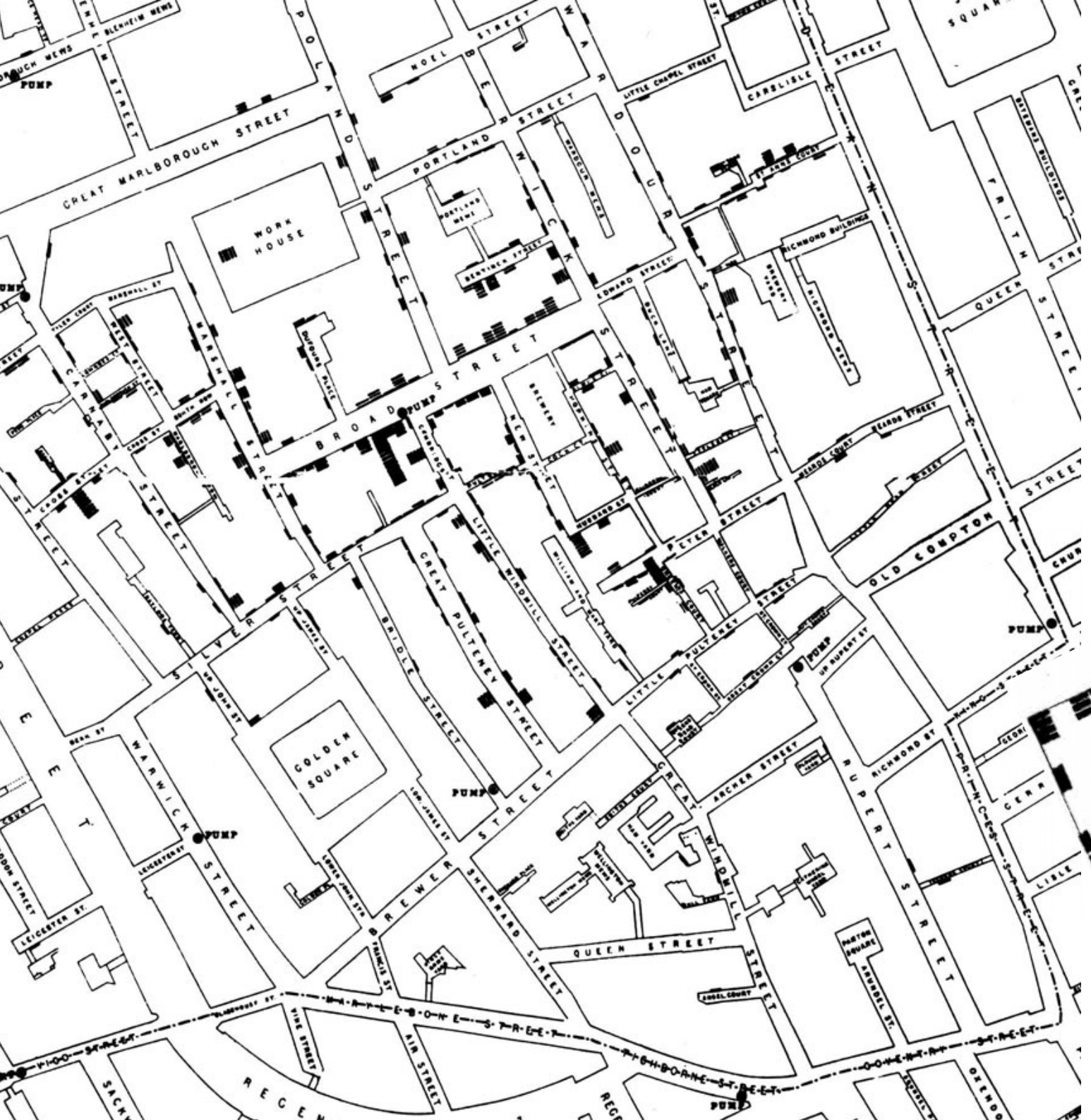
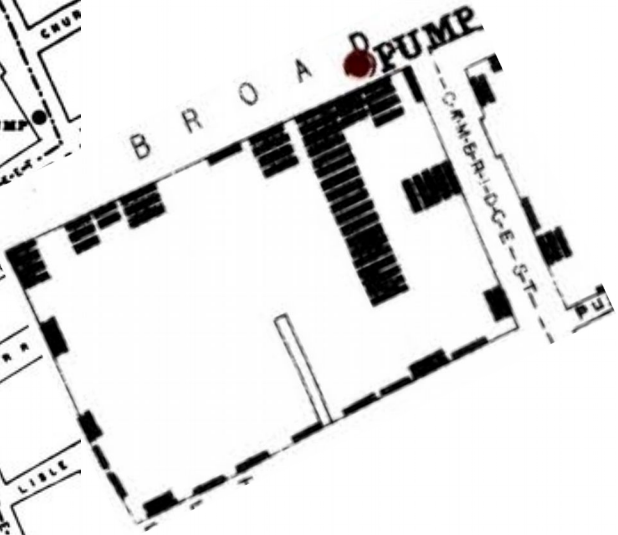


FIG. 8. *The bigger-means-more metaphor of this dot-array map affords a more appropriate treatment of the count data in Figure 7.*





John Snow 1854.

Depende de cómo dibujes los límites.

How to lie with maps. Monmonier

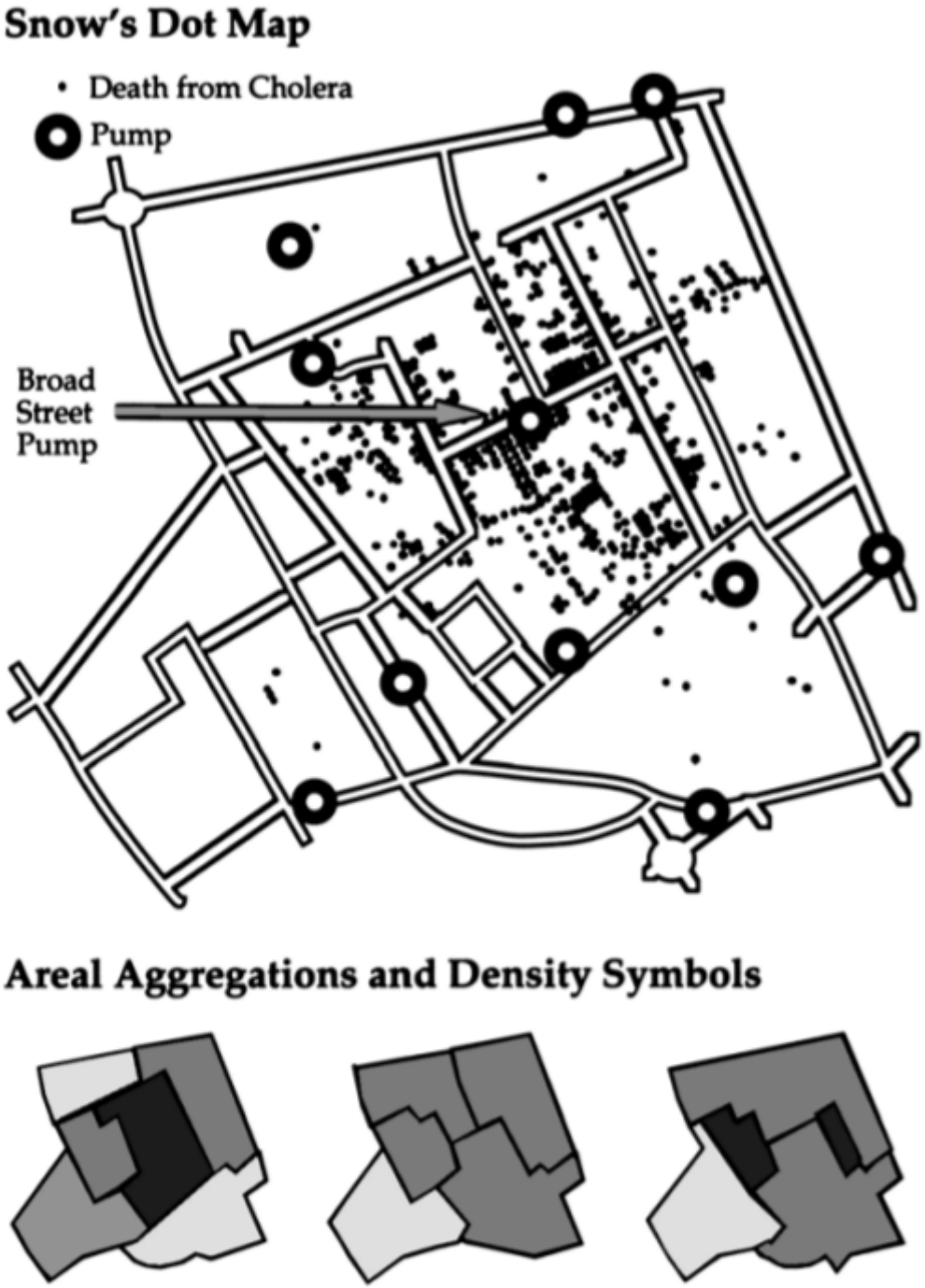
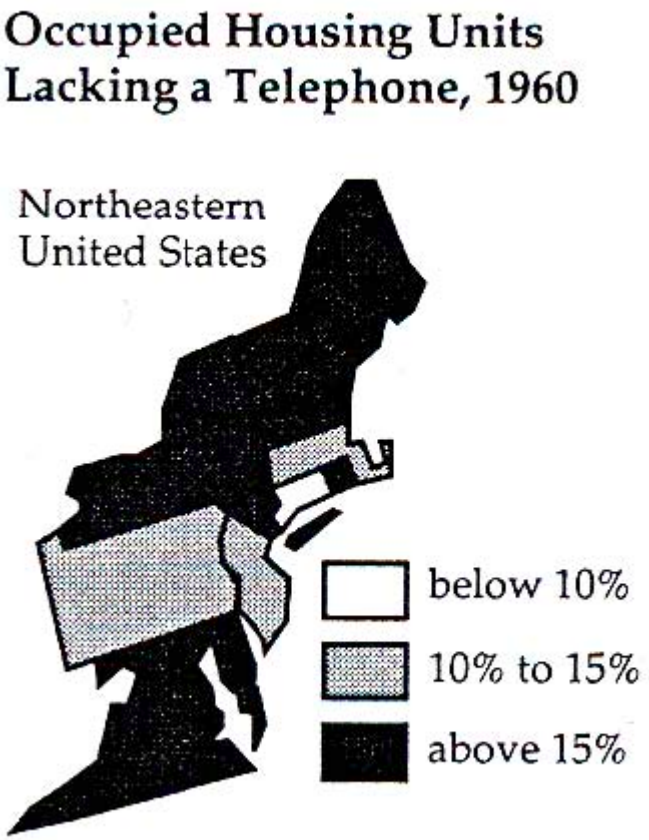
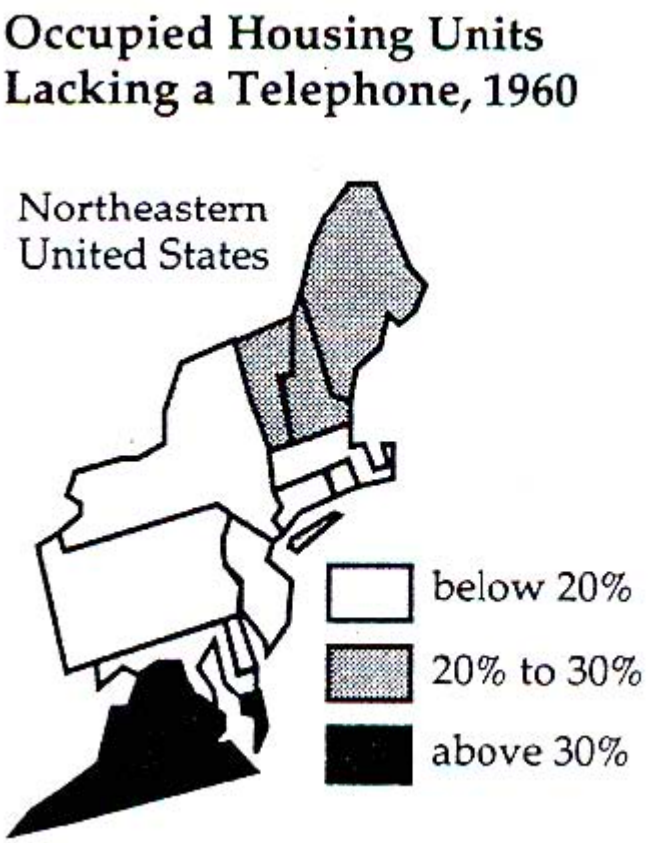
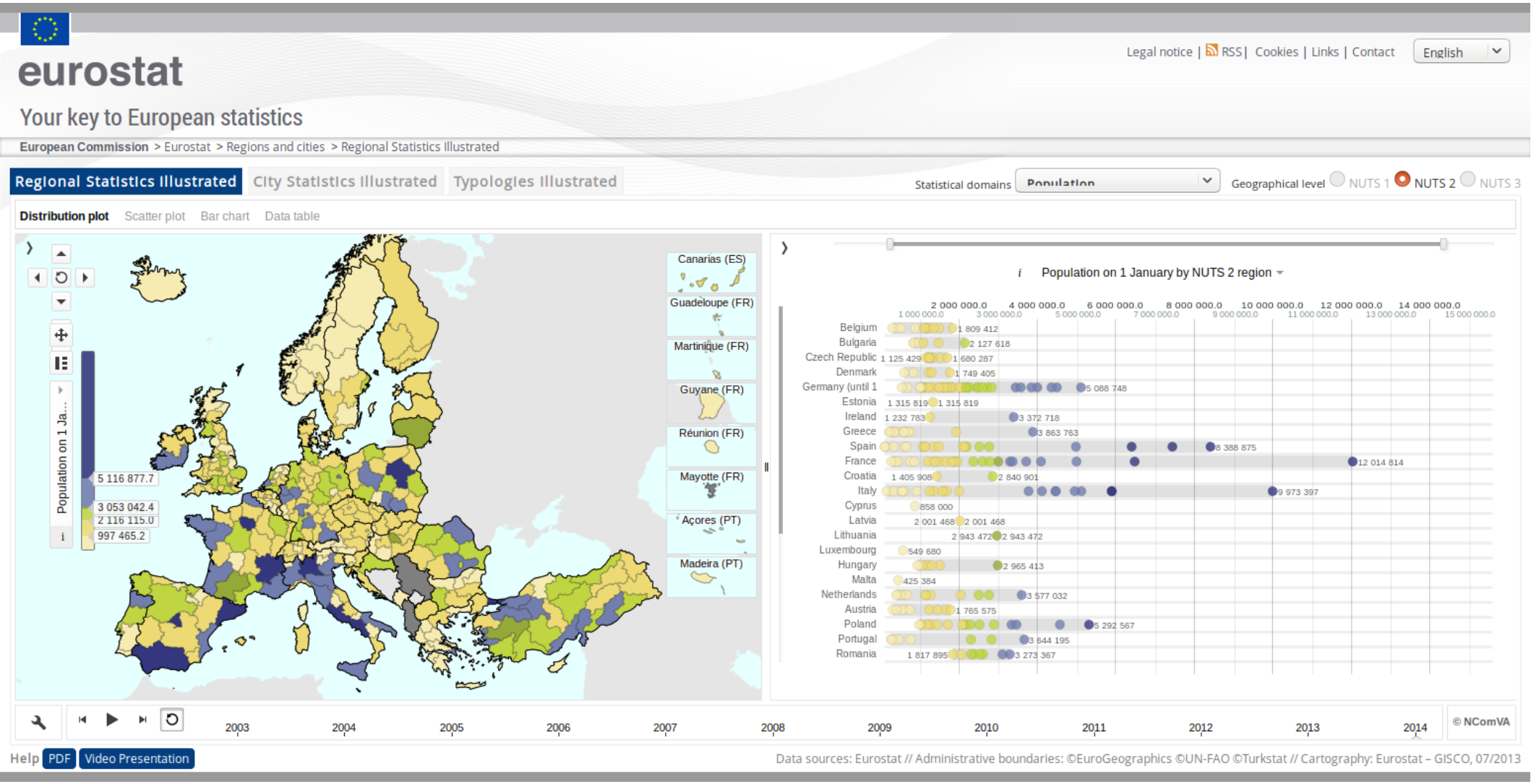


FIGURE 10.18. A reconstruction of John Snow's famous dot map of cholera (above) and three choropleth maps (below) produced by different areal aggregations of this part of London.





Depende de cómo colorees el mapa: colores y rangos.



Eruostat: data and distibution plot. Population.

http://ec.europa.eu/eurostat/cache/RSI/#?vis=nuts2.population&lang=en

Hace un año...

Google Maps it not The Map
Taller de cartografías digitales ciudadanas

29-31 abril 2016. Hirikilabs. Tabakalera. Donosti.

Info sobre el mapa

Foto aérea

Mapa

Vectores

Datos

Interfaz de usuario

genera

Info sobre
el mapa

Foto aérea

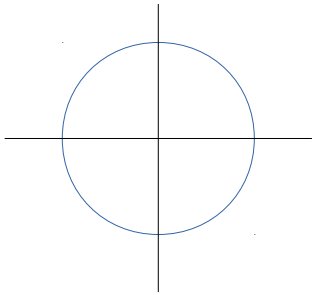
Mapa

Vectores
Datos

Interfaz
de usuario

genera

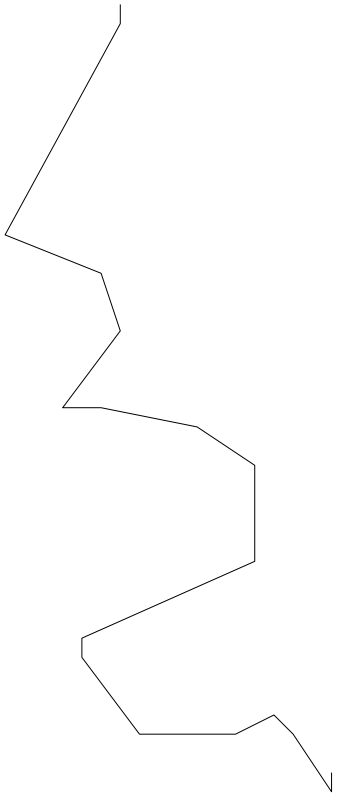
Punto



Posición:
- Latitud
- Longitud

Tiempo
Elevación
Nombre
Precisión GPS
Categoría
(Velocidad)

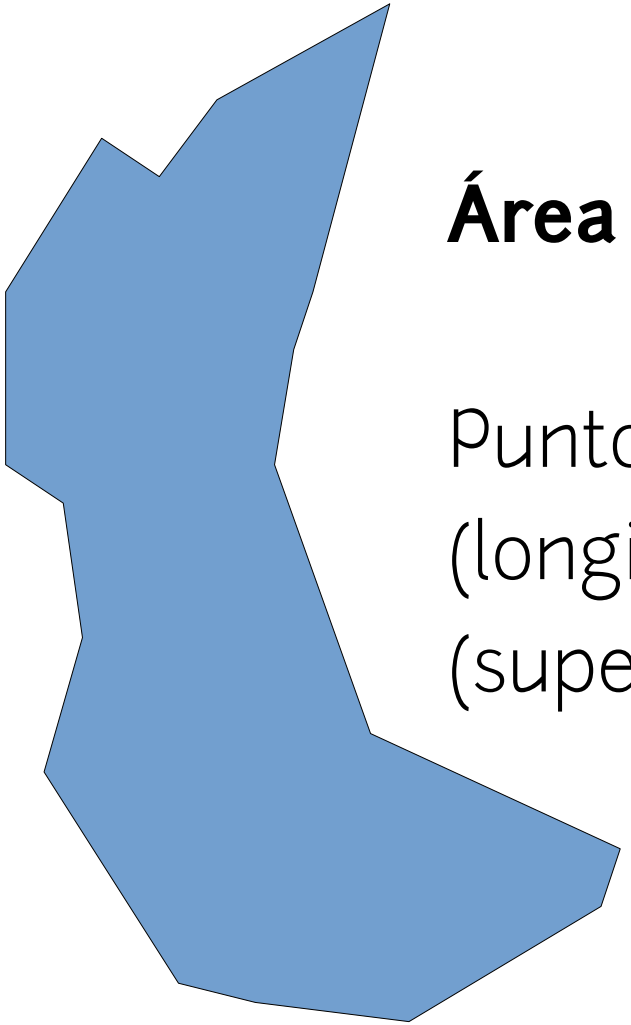
.csv
.tsv
.geojson
.kml
.gpx



Línea

Puntos
(longitud)

.csv
.tsv
.geojson
.kml
.gpx

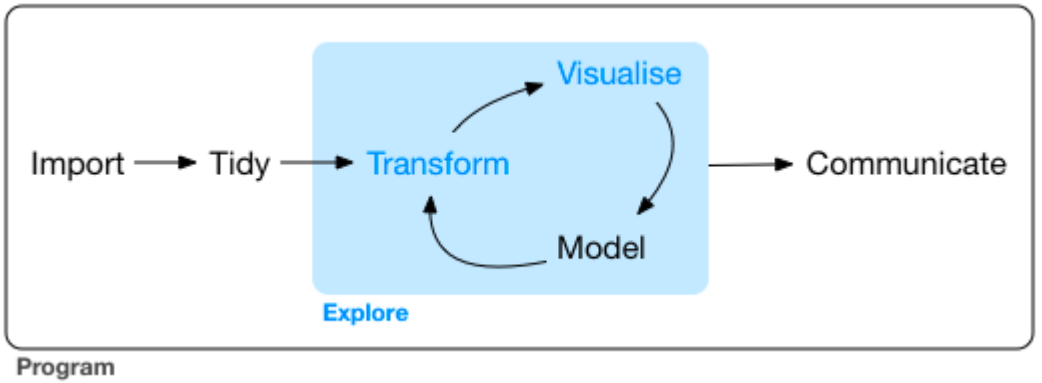


Área

Puntos
(longitud)
(superficie)

.csv
.tsv
.geojson
.kml
.gpx
shapefile

Proceso de visualización de datos



1. Obtención
2. Limpieza
3. Preparación
4. Exploración y análisis
5. Publicación

¿Cómo vamos a trabajar?

Herramientas abiertas y/o de software libre.

Proceso semidirigido.

Artesanal / semiautomatizado. Ejemplos

Manos a la obra

Documentos compartido

http://piratepad.net/airbnbeuskadi

¿Cómo vamos a trabajar?

Herramientas abiertas y/o de software libre.

Proceso semidirigido.

Artesanal / semiautomatizado.

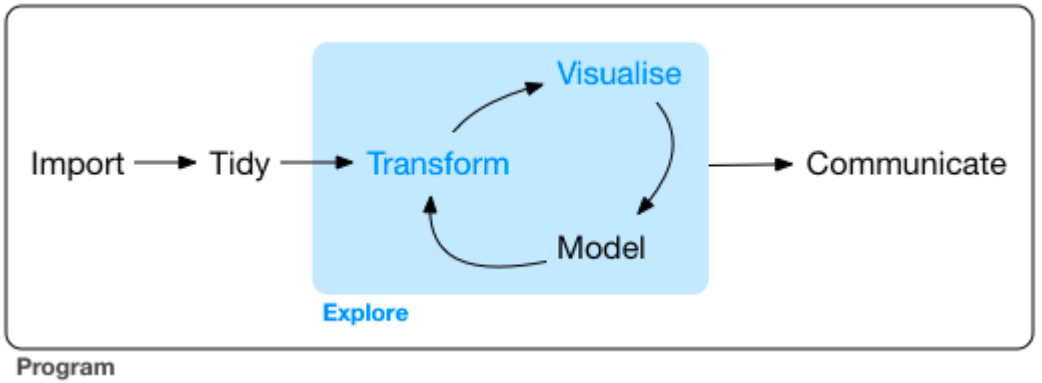
–

Un paseo por:

Github.com

Piratepad.net

Proceso de visualización de datos



0. Atención: pregunta
1. Obtención (Inside Airbnb, Open DataEuskadi, GeoEuskadi, Euskalgeo)
2. Limpieza (Open Refine)
3. Preparación y transformación (R, QGIS)
4. Exploración (carto, R, LibreOffice, google spreadsheet)
5. Publicación (html)

1. Obtención de datos

1. Fuentes oficiales
2. Fuentes semioficiales
3. Estudios / papers

...o scraping

Inside Airbnb: fuente principal de datos. Web scraping por Murray Cox.

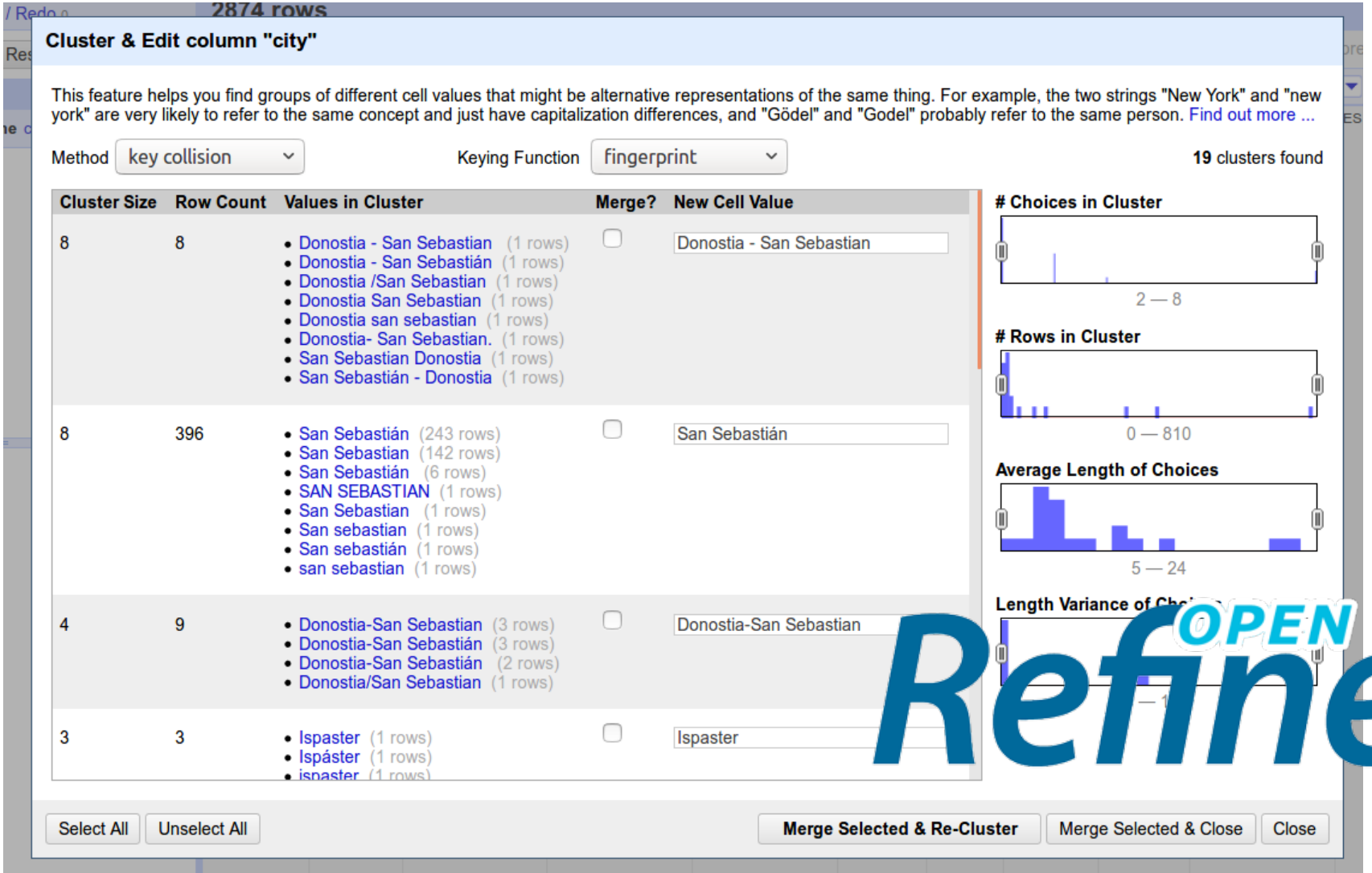
Open DataEuskadi: número de viviendas por barrio, alojamientos turísticos.

GeoEuskadi: Contornos de barrios.

Euskalgeo: ayuntamientos.

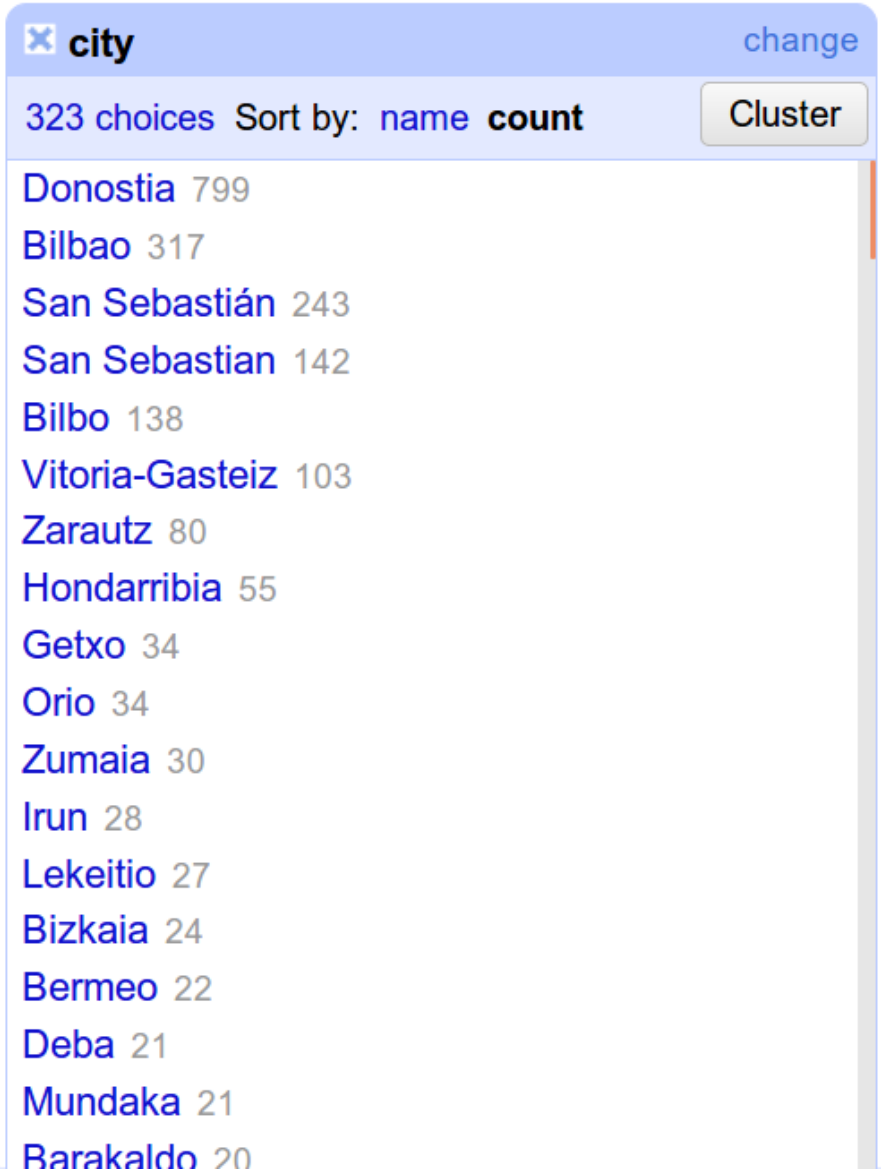
2. Limpieza

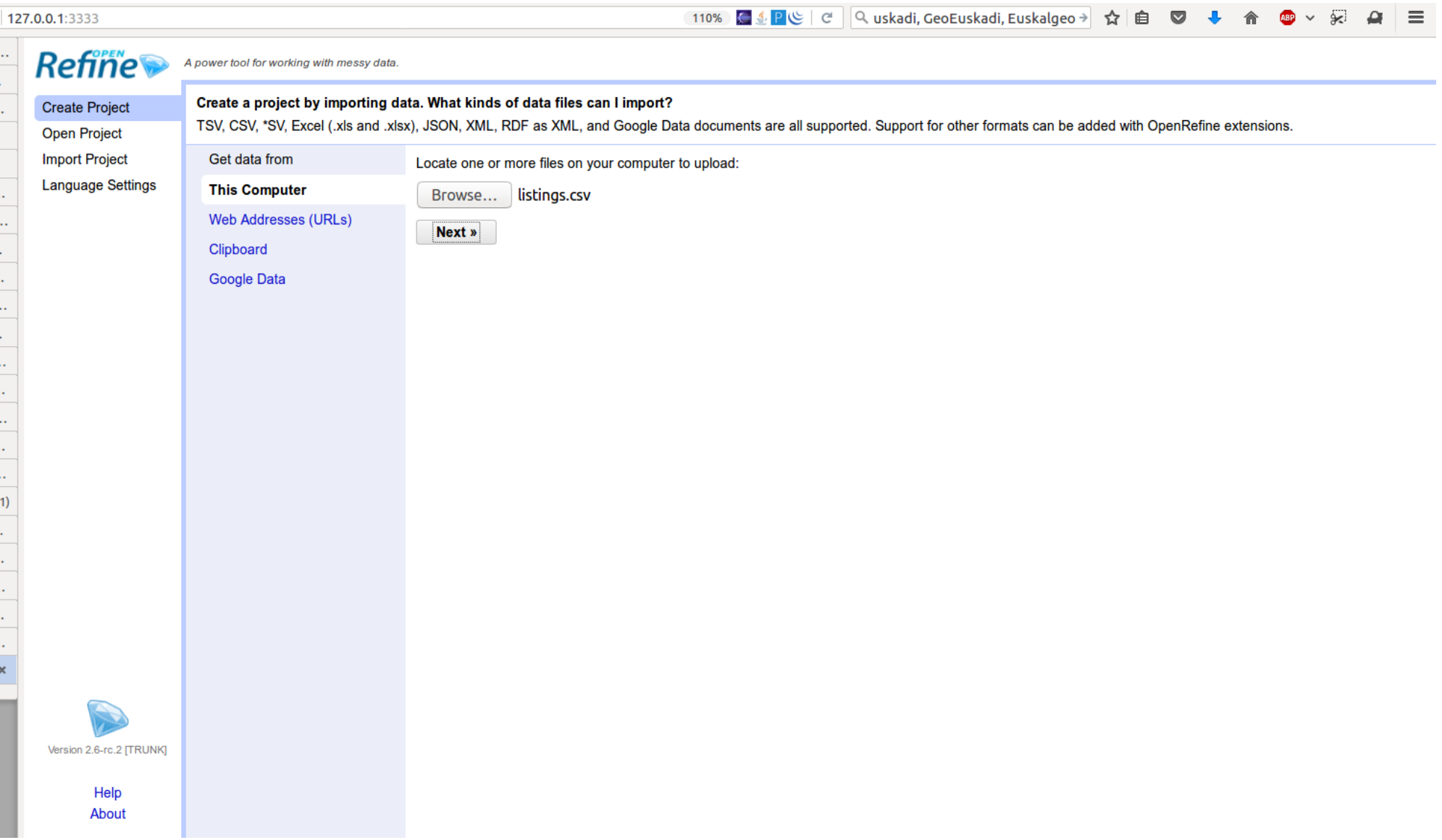
Si tus datos están sucios puedes limpiarlos:
Open Refine (http://openrefine.org/)

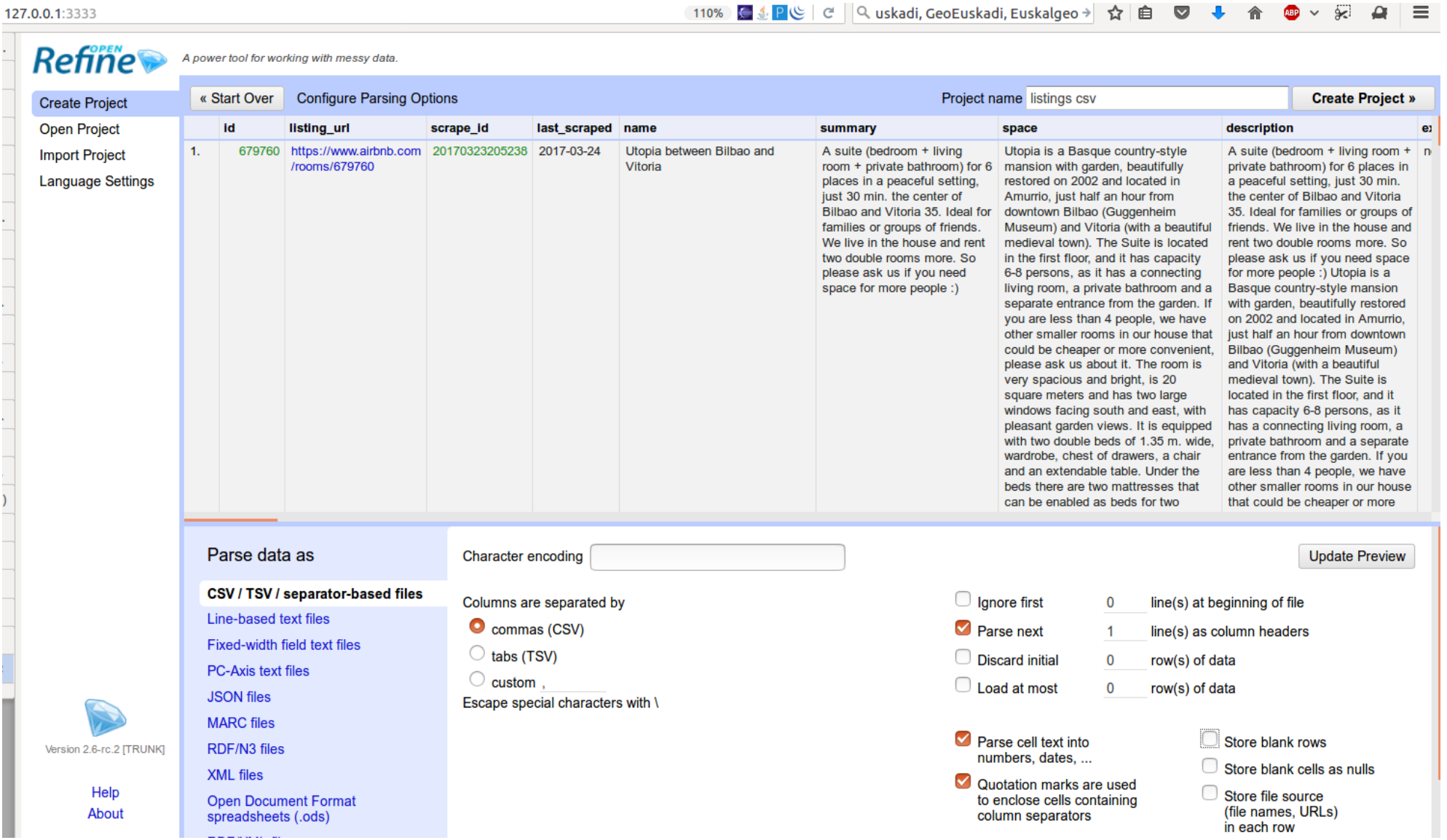


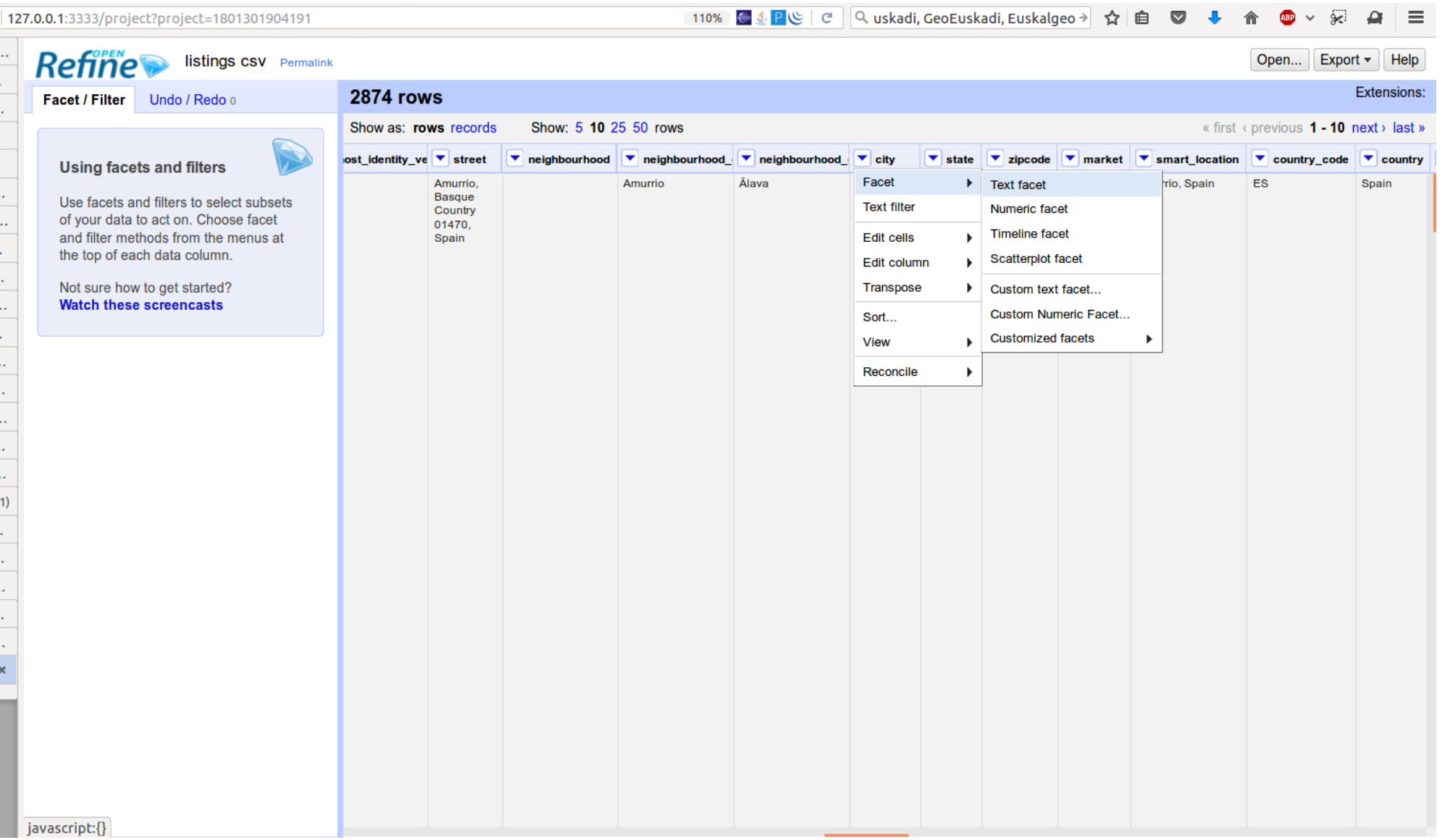
Limpiando con Open Refine

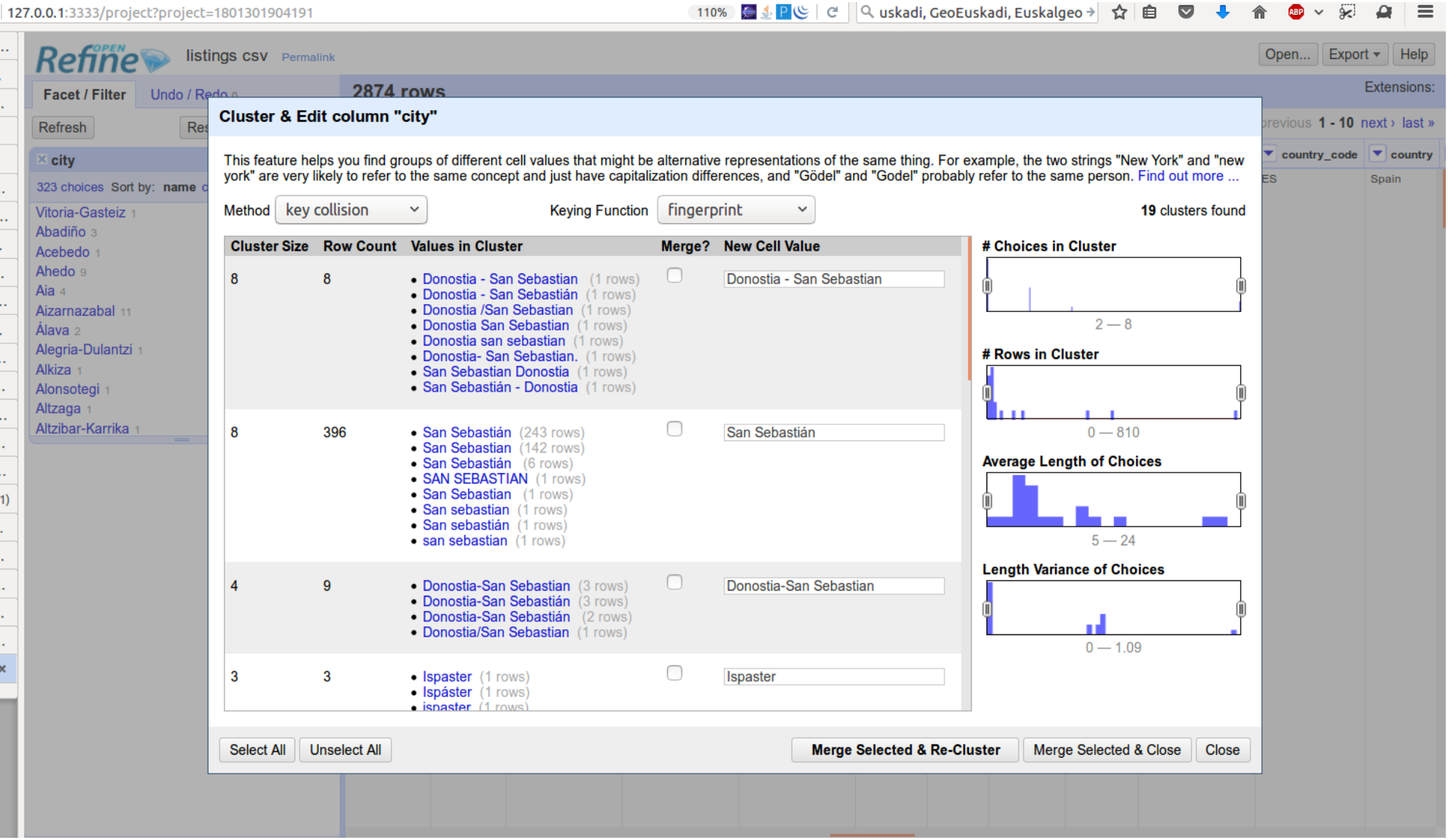
1. Descarga repositorio https://github.com/montera34/airbnbeuskadi/
2. Descomprime archivo **/data/original/listings_euskadi.csv.gz**
3. Descarga Open Refine http://openrefine.org/download.html
4. Instala Open Refine
5. Limpia, fija y da esplendor.

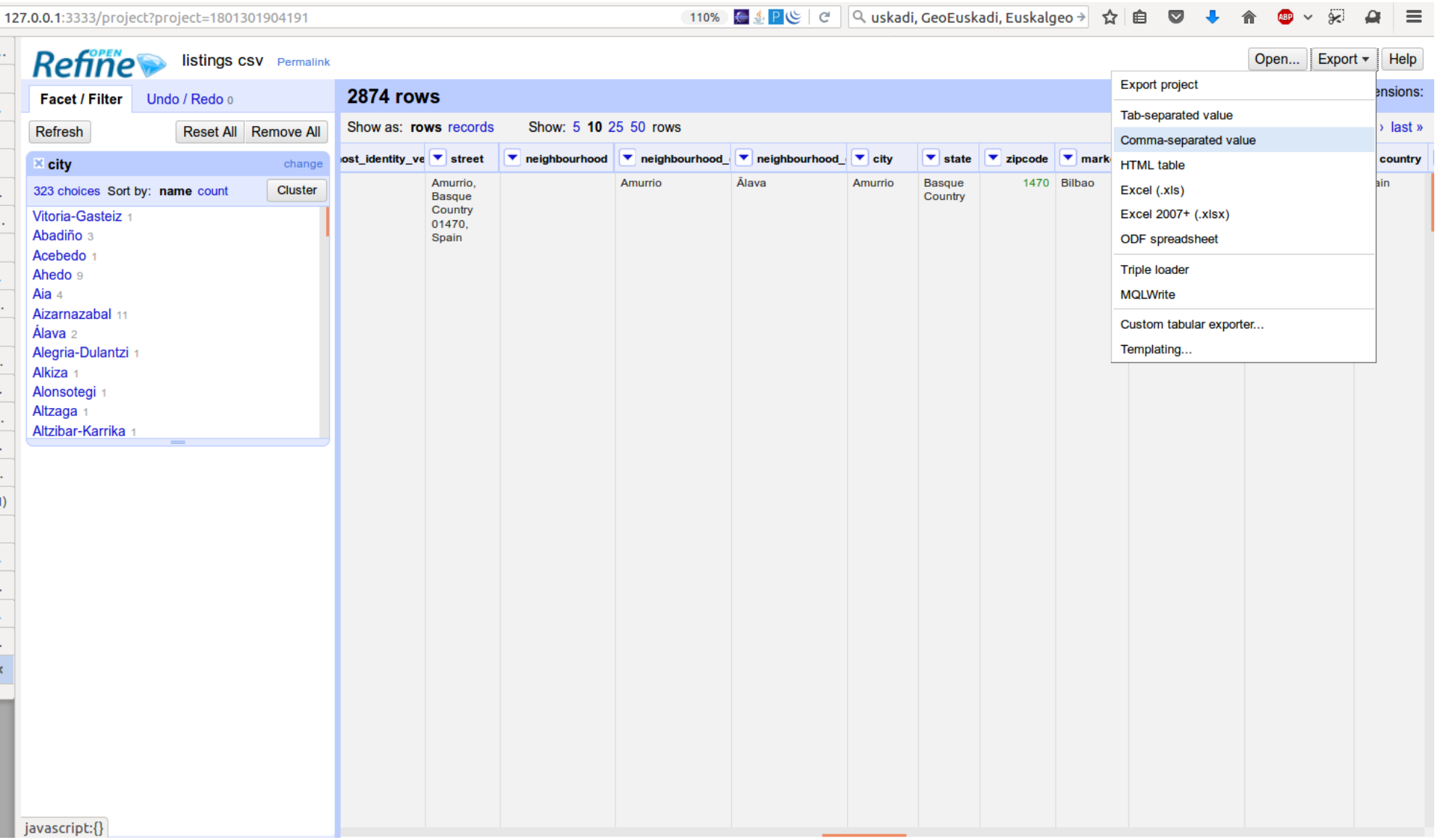

city
change
323 choices Sort by: name count
Cluster
Donostia 799
Bilbao 317
San Sebastián 243
San Sebastian 142
Bilbo 138
Vitoria-Gasteiz 103
Zarautz 80
Hondarribia 55
Getxo 34
Orio 34
Zumaia 30
Irun 28
Lekeitio 27
Bizkaia 24
Bermeo 22
Deba 21
Mundaka 21
Barakaldo 20


127.0.0.1:3333
110%
uskadi, GeoEuskadi, Euskalgeo
OpenRefine
A power tool for working with messy data.
Create Project
Open Project
Import Project
Language Settings
Create a project by importing data. What kinds of data files can I import?
TSV, CSV, *SV, Excel (.xls and .xlsx), JSON, XML, RDF as XML, and Google Data documents are all supported. Support for other formats can be added with OpenRefine extensions.
Get data from
This Computer
Web Addresses (URLs)
Clipboard
Google Data
Locate one or more files on your computer to upload:
Browse...
listings.csv
Next »
Version 2.6-rc.2 [TRUNK]
Help
About


127.0.0.1:3333
110%
uskadi, GeoEuskadi, Euskalgeo
OpenRefine
A power tool for working with messy data.
Create Project
Open Project
Import Project
Language Settings
« Start Over
Configure Parsing Options
Project name
listings csv
Create Project »
id
listing_url
scrape_id
last_scraped
name
summary
space
description
1.
679760
https://www.airbnb.com/rooms/679760
20170323205238
2017-03-24
Utopia between Bilbao and Vitoria
A suite (bedroom + living room + private bathroom) for 6 places in a peaceful setting, just 30 min. the center of Bilbao and Vitoria 35. Ideal for families or groups of friends. We live in the house and rent two double rooms more. So please ask us if you need space for more people :)
Utopia is a Basque country-style mansion with garden, beautifully restored on 2002 and located in Amurrio, just half an hour from downtown Bilbao (Guggenheim Museum) and Vitoria (with a beautiful medieval town). The Suite is located in the first floor, and it has capacity 6-8 persons, as it has a connecting living room, a private bathroom and a separate entrance from the garden. If you are less than 4 people, we have other smaller rooms in our house that could be cheaper or more convenient, please ask us about it. The room is very spacious and bright, is 20 square meters and has two large windows facing south and east, with pleasant garden views. It is equipped with two double beds of 1.35 m. wide, wardrobe, chest of drawers, a chair and an extendable table. Under the beds there are two mattresses that can be enabled as beds for two
A suite (bedroom + living room + private bathroom) for 6 places in a peaceful setting, just 30 min. the center of Bilbao and Vitoria 35. Ideal for families or groups of friends. We live in the house and rent two double rooms more. So please ask us if you need space for more people :) Utopia is a Basque country-style mansion with garden, beautifully restored on 2002 and located in Amurrio, just half an hour from downtown Bilbao (Guggenheim Museum) and Vitoria (with a beautiful medieval town). The Suite is located in the first floor, and it has capacity 6-8 persons, as it has a connecting living room, a private bathroom and a separate entrance from the garden. If you are less than 4 people, we have other smaller rooms in our house that could be cheaper or more
Parse data as
CSV / TSV / separator-based files
Line-based text files
Fixed-width field text files
PC-Axis text files
JSON files
MARC files
RDF/N3 files
XML files
Open Document Format spreadsheets (.ods)
Character encoding
Update Preview
Columns are separated by
commas (CSV)
tabs (TSV)
custom ,
Escape special characters with \
Ignore first 0 line(s) at beginning of file
Parse next 1 line(s) as column headers
Discard initial 0 row(s) of data
Load at most 0 row(s) of data
Parse cell text into numbers, dates, ...
Quotation marks are used to enclose cells containing column separators
Store blank rows
Store blank cells as nulls
Store file source (file names, URLs) in each row
Version 2.6-rc.2 [TRUNK]
Help
About


127.0.0.1:3333/project?project=1801301904191
110%
uskadi, GeoEuskadi, Euskalgeo
OpenRefine
listings csv
Permalink
Open...
Export
Help
Facet / Filter
Undo / Redo 0
Using facets and filters
Use facets and filters to select subsets of your data to act on. Choose facet and filter methods from the menus at the top of each data column.
Not sure how to get started?
Watch these screencasts
2874 rows
Extensions:
Show as: rows records
Show: 5 10 25 50 rows
« first ‹ previous 1 - 10 next › last »
ost_identity_ve
street
neighbourhood
neighbourhood_
neighbourhood_
city
state
zipcode
market
smart_location
country_code
country
Amurrio, Basque Country 01470, Spain
Amurrio
Álava
rio, Spain
ES
Spain
Facet
Text filter
Edit cells
Edit column
Transpose
Sort...
View
Reconcile
Text facet
Numeric facet
Timeline facet
Scatterplot facet
Custom text facet...
Custom Numeric Facet...
Customized facets
javascript:{}


127.0.0.1:3333/project?project=1801301904191
OpenRefine
listings csv
Permalink
Open...
Export
Help
Facet / Filter
Undo / Redo
2874 rows
Extensions:
Refresh
city
323 choices Sort by: name
Vitoria-Gasteiz 1
Abadiño 3
Acebedo 1
Ahedo 9
Aia 4
Aizarnazabal 11
Álava 2
Alegria-Dulantzi 1
Alkiza 1
Alonsotegi 1
Altzaga 1
Altzibar-Karrika 1
previous 1 - 10 next › last »
country_code
country
ES
Spain
Cluster & Edit column "city"
This feature helps you find groups of different cell values that might be alternative representations of the same thing. For example, the two strings "New York" and "new york" are very likely to refer to the same concept and just have capitalization differences, and "Gödel" and "Godel" probably refer to the same person. Find out more ...
Method key collision
Keying Function fingerprint
19 clusters found
Cluster Size
Row Count
Values in Cluster
Merge?
New Cell Value
8
8
• Donostia - San Sebastian (1 rows)
• Donostia - San Sebastián (1 rows)
• Donostia /San Sebastian (1 rows)
• Donostia San Sebastian (1 rows)
• Donostia san sebastian (1 rows)
• Donostia- San Sebastian. (1 rows)
• San Sebastian Donostia (1 rows)
• San Sebastián - Donostia (1 rows)
Donostia - San Sebastian
8
396
• San Sebastián (243 rows)
• San Sebastian (142 rows)
• San Sebastián (6 rows)
• SAN SEBASTIAN (1 rows)
• San Sebastian (1 rows)
• San sebastian (1 rows)
• San sebastián (1 rows)
• san sebastian (1 rows)
San Sebastián
4
9
• Donostia-San Sebastian (3 rows)
• Donostia-San Sebastián (3 rows)
• Donostia-San Sebastián (2 rows)
• Donostia/San Sebastian (1 rows)
Donostia-San Sebastian
3
3
• Ispaster (1 rows)
• Ispáster (1 rows)
• ispaster (1 rows)
Ispaster
Choices in Cluster
2 — 8
Rows in Cluster
0 — 810
Average Length of Choices
5 — 24
Length Variance of Choices
0 — 1.09
Select All
Unselect All
Merge Selected & Re-Cluster
Merge Selected & Close
Close

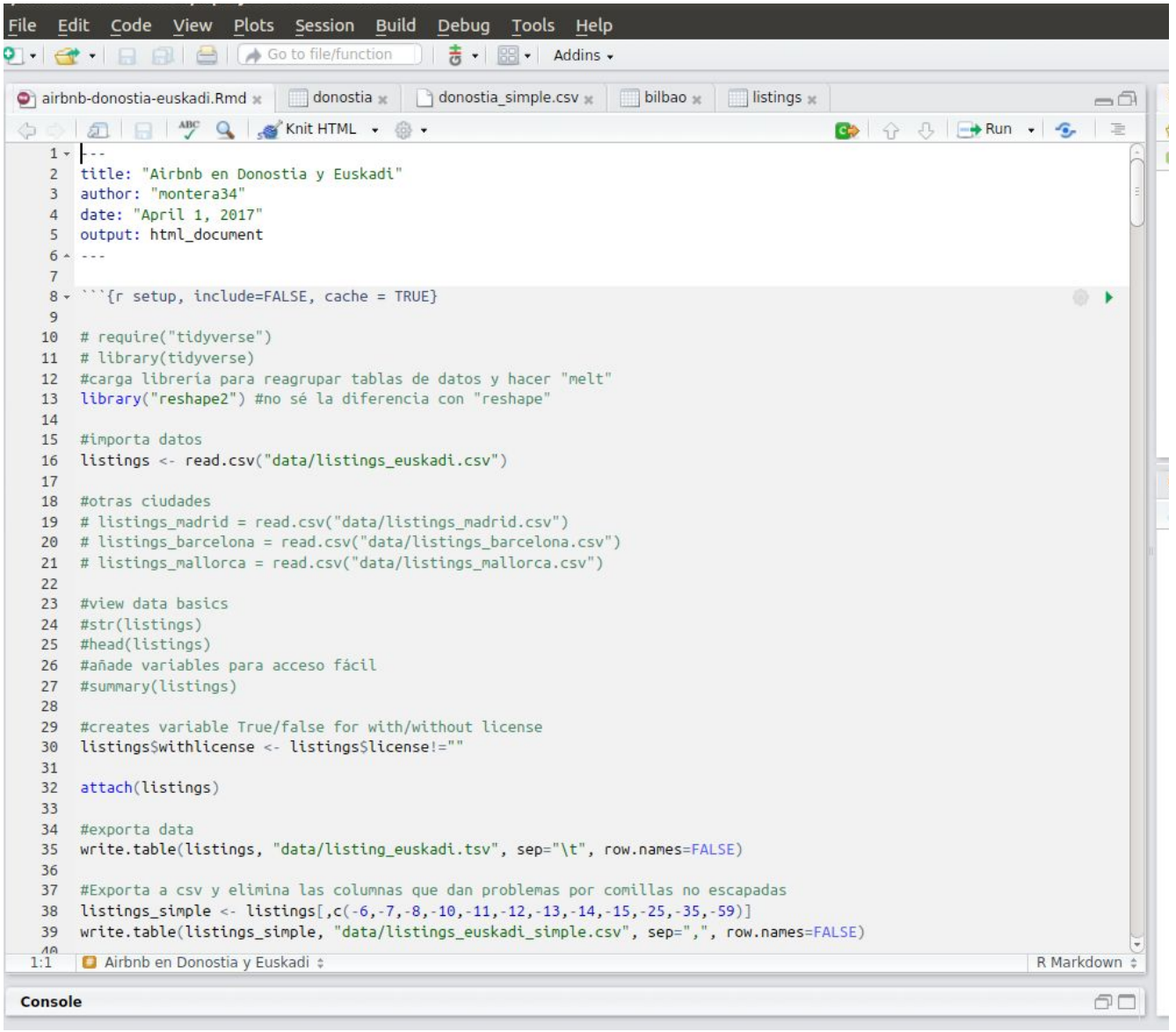

127.0.0.1:3333/project?project=1801301904191
110%
uskadi, GeoEuskadi, Euskalgeo
OpenRefine
listings csv
Permalink
Open...
Export
Help
Facet / Filter
Undo / Redo 0
Refresh
Reset All
Remove All
city
change
323 choices Sort by: name count
Cluster
Vitoria-Gasteiz 1
Abadiño 3
Acebedo 1
Ahedo 9
Aia 4
Aizarnazabal 11
Álava 2
Alegria-Dulantzi 1
Alkiza 1
Alonsotegi 1
Altzaga 1
Altzibar-Karrika 1
2874 rows
Show as: rows records
Show: 5 10 25 50 rows
› last »
ost_identity_ve
street
neighbourhood
neighbourhood_
neighbourhood_
city
state
zipcode
mark
country
Amurrio, Basque Country 01470, Spain
Amurrio
Álava
Amurrio
Basque Country
1470
Bilbao
Export project
Tab-separated value
Comma-separated value
HTML table
Excel (.xls)
Excel 2007+ (.xlsx)
ODF spreadsheet
Triple loader
MQLWrite
Custom tabular exporter...
Templating...
javascript:{}

3. Preparación y transformación

- Cambiar sistema de proyección en mapas (QGIS)
- Limpiar dobles comillas (a mano)
- Quitar símbolos de dolar ($) de los precios (busca y reemplaza)
- Añadir nuevas variables (R)
- Simplificar: bases de datos con menos campos (R)
- Simplificar: bases de datos de cada ciudad (R)

Un TV 42" puede estropearte la tarde.

Simplificando con R



Simplificando con R

Generar los archivos por ciudad y simplificados:

- listings_bilbao.csv
- listings_bilbao_simple.csv
- listings_donostia.csv
- listings_donostia_simple.csv
- listings_euskadi.csv
- listings_euskadi_simple.csv
- listings_vitoria.csv
- listings_vitoria_simple.csv

Simplificando con R

Generar los archivos por ciudad y simplificados (Rmarkdown):

Airbnb en Euskadi

La base de datos de Euskadi de InsideAirbnb contiene 2874 alojamientos en 267 municipios, de ellos 395 tienen licencia de alojamiento turístico (13.7%).

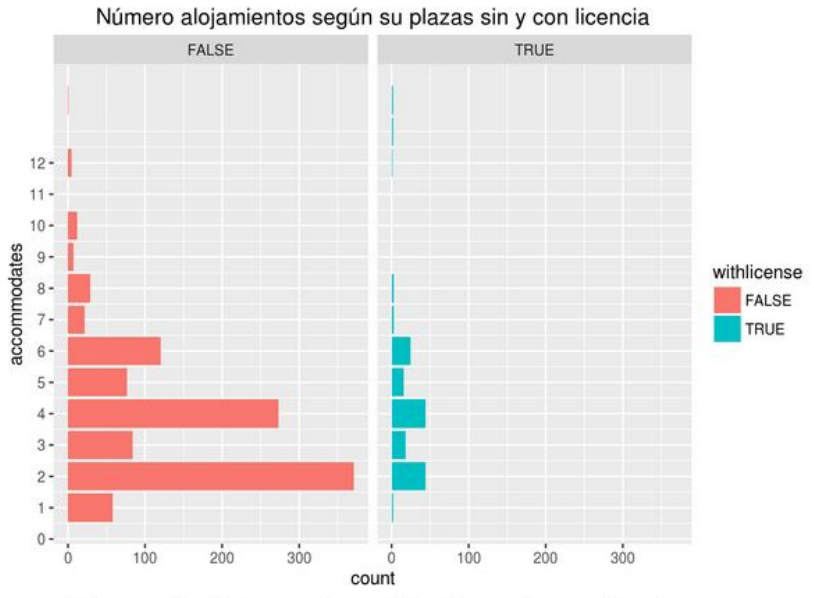
En las capitales tenemos:

- **Donostia - San Sebastián** 1219 (42.4 % del total), de esos alojamientos tienen licencia 160 (13.1 % del municipio). Dispone de 4553 plazas. Usuarios con alojamiento listado 696
- **Bilbao** 459 (16.0% del total), de esos alojamientos tienen licencia 74 (16.1 % del municipio). Dispone de 1484 plazas. Usuarios con alojamiento listado 367
- **Vitoria - Gasteiz** 107 (3.7% del total), de esos alojamientos tienen licencia 27 (25.2 % del municipio). Dispone de 361 plazas. Usuarios con alojamiento listado 79

Por comparar con otras ciudades, el porcentaje de alojamientos con licencia turística en Barcelona es 18.0% de un total de 17369 alojamientos según datos de insideairbnb. Madrid (7446 alojamientos) y Mallorca (11271 alojamientos) no tienen indicado si tienen licencia.

Donostia - San Sebastián

En Donostia existen 696 usuarios con alojamientos disponibles en Airbnb.



https://numeroteca.github.io/airbnb-donostia-euskadi.html

4. Exploración: visualizar, transformar, probar y vuelta a empezar

> La exploración de los datos es el arte de mirar a tus datos y rápidamente generar hipótesis, probarlas inmediatamente y repetir una y otra vez una y otra vez. El objetivo de la exploración de datos es generar ideas prometedoras que luego podrás explorar más tarde en profundidad.
>
> R for data Science. http://r4ds.had.co.nz/explore-intro.html

4. Exploración: Pivot tables (tablas dinámicas)

Archivo: data/listings_donostia_simple.csv

Herramientas: Google spreadsheet o LibreOffice.

Objetivo: Creación de pivot tables para entender los datos. Obtener números gordos.

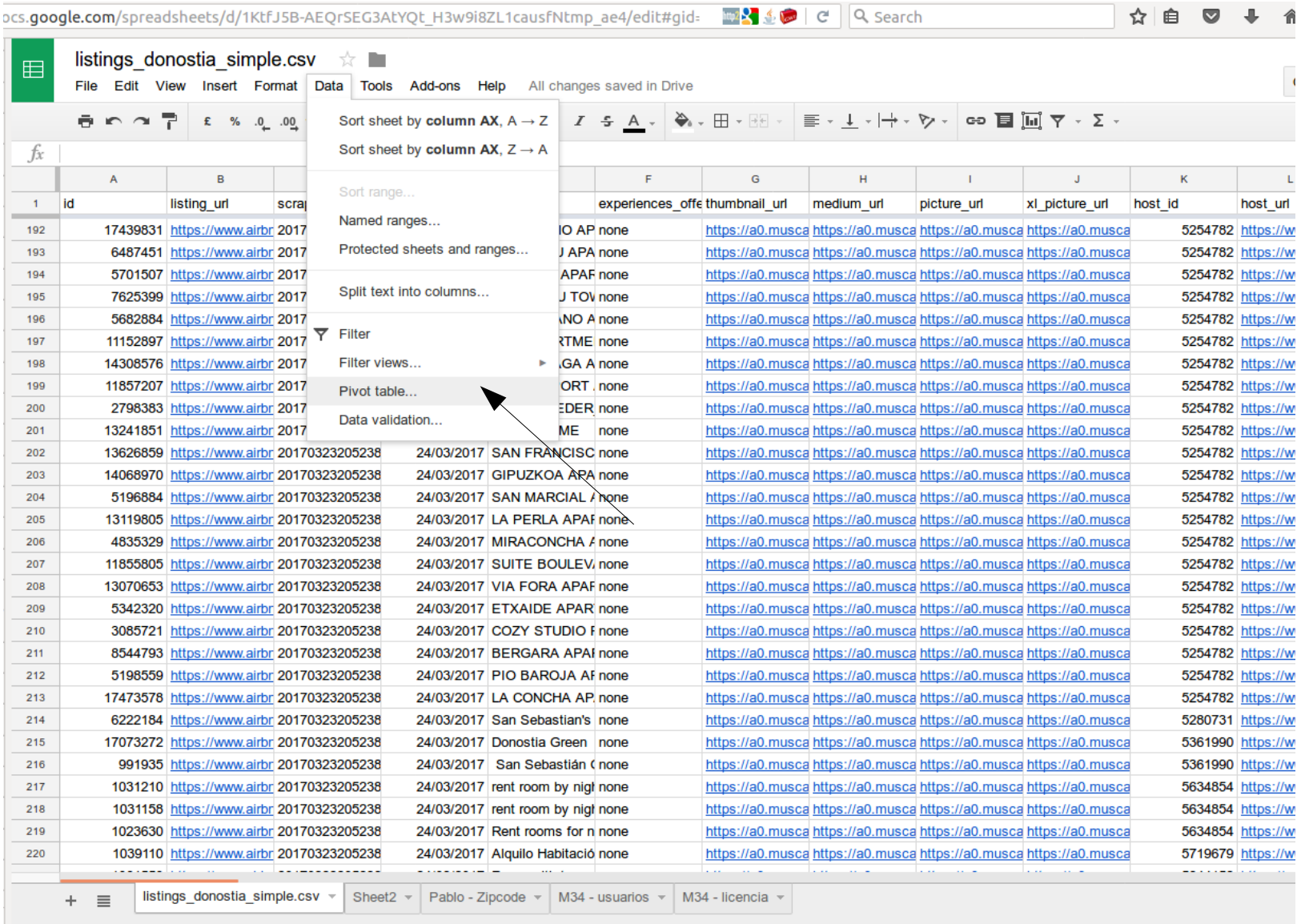
Documento: https://docs.google.com/spreadsheets/d/1KtfJ5B-AEQrSEG3AtYQt_H3wgi8ZL1causfNtmp_ae4

Ejercicio. Número y % de alojamientos:

- Por usuario
- Por tipo de alojamiento
- Con/sin licencia turística
- Número de plazas
- ...

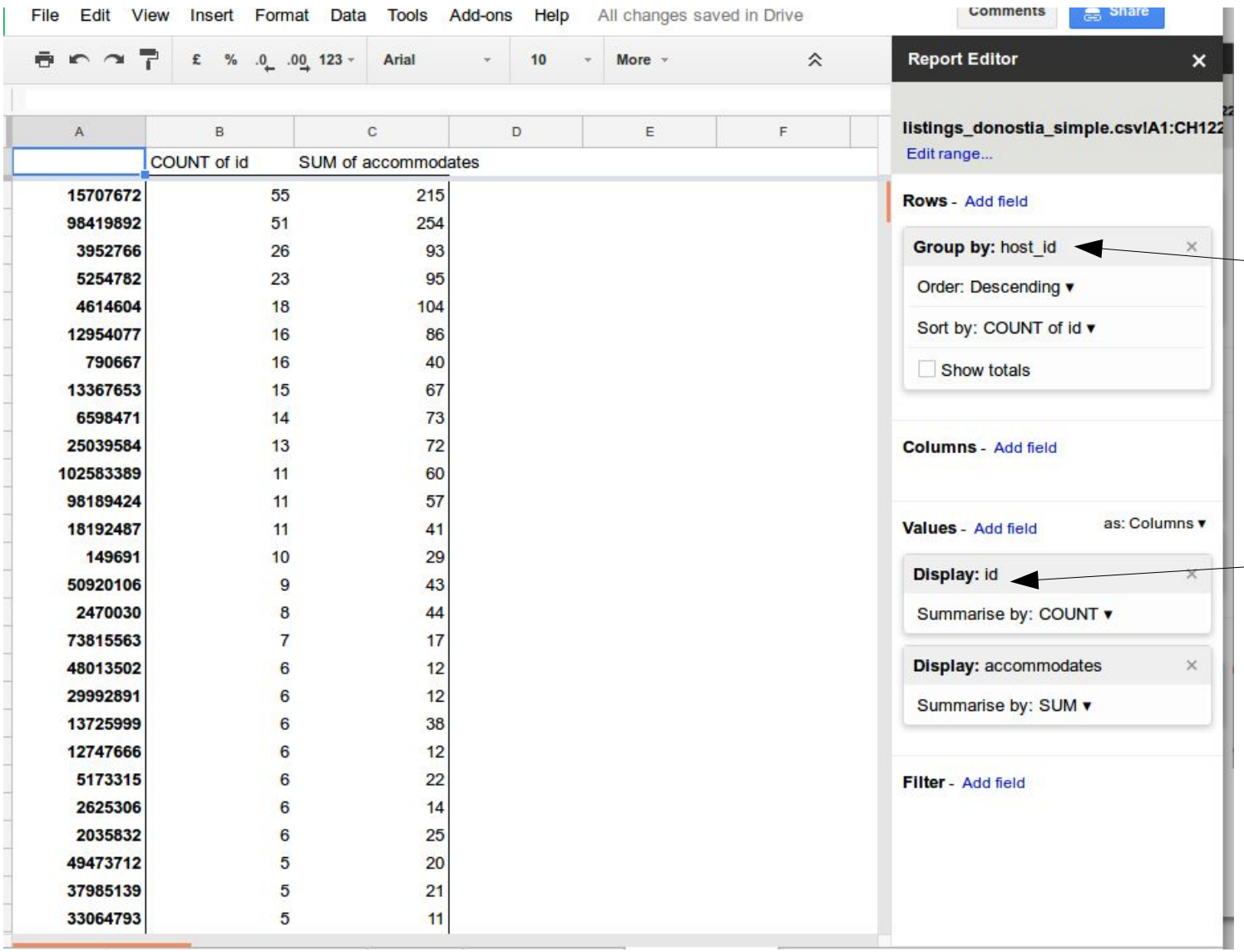
Selecciona tu variable en el doc compartido.

Crea pivot table: Data > Pivot table



https://docs.google.com/spreadsheets/d/1KtfJ5B-AEQrSEG3AtYQt_H3w9i8ZL1causfNtmp_ae4/

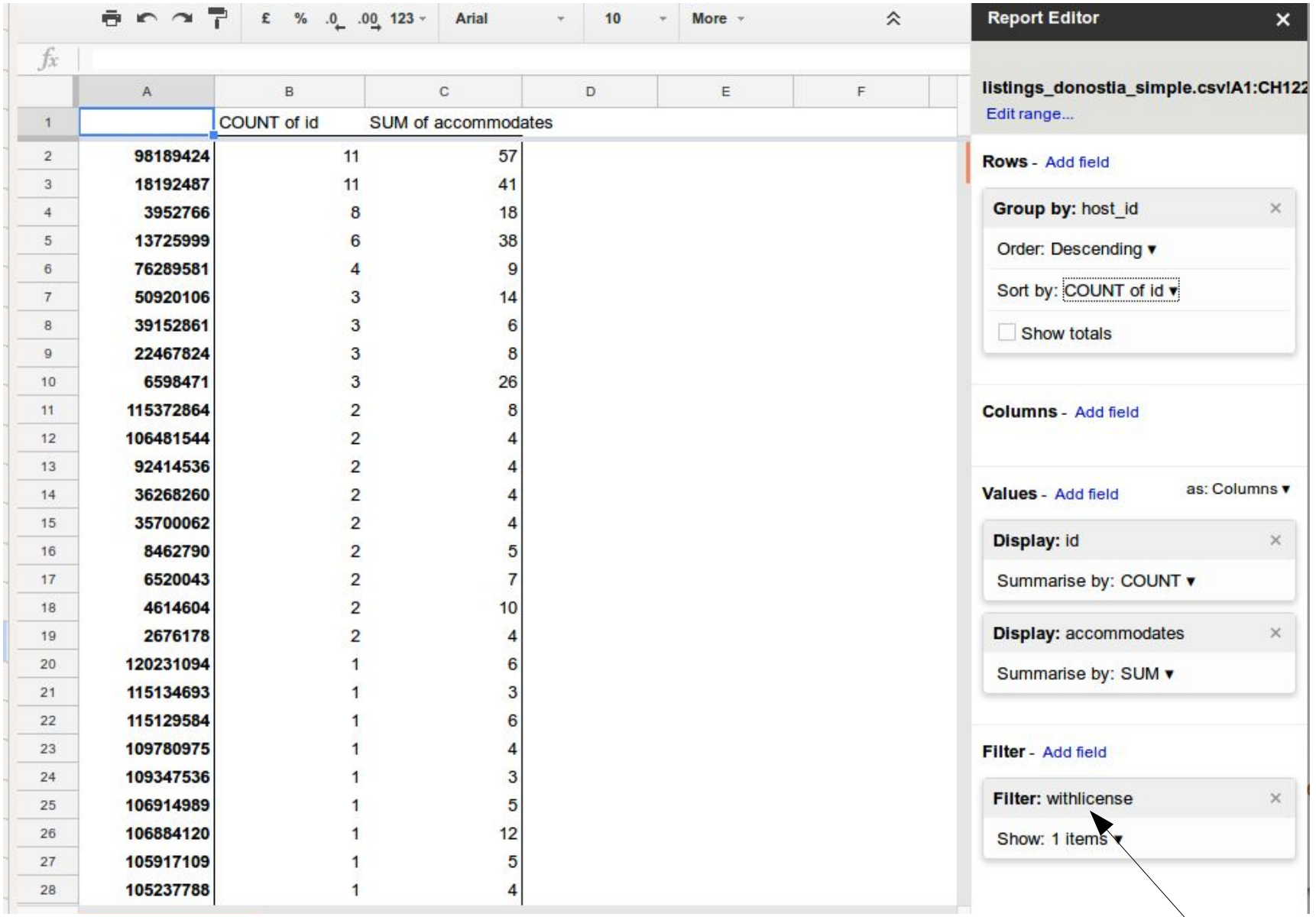
Edita pivto table



	COUNT of id	SUM of accommodates
15707672	55	215
98419892	51	254
3952766	26	93
5254782	23	95
4614604	18	104
12954077	16	86
790667	16	40
13367653	15	67
6598471	14	73
25039584	13	72
102583389	11	60
98189424	11	57
18192487	11	41
149691	10	29
50920106	9	43
2470030	8	44
73815563	7	17
48013502	6	12
29992891	6	12
13725999	6	38
12747666	6	12
5173315	6	22
2625306	6	14
2035832	6	25
49473712	5	20
37985139	5	21
33064793	5	11

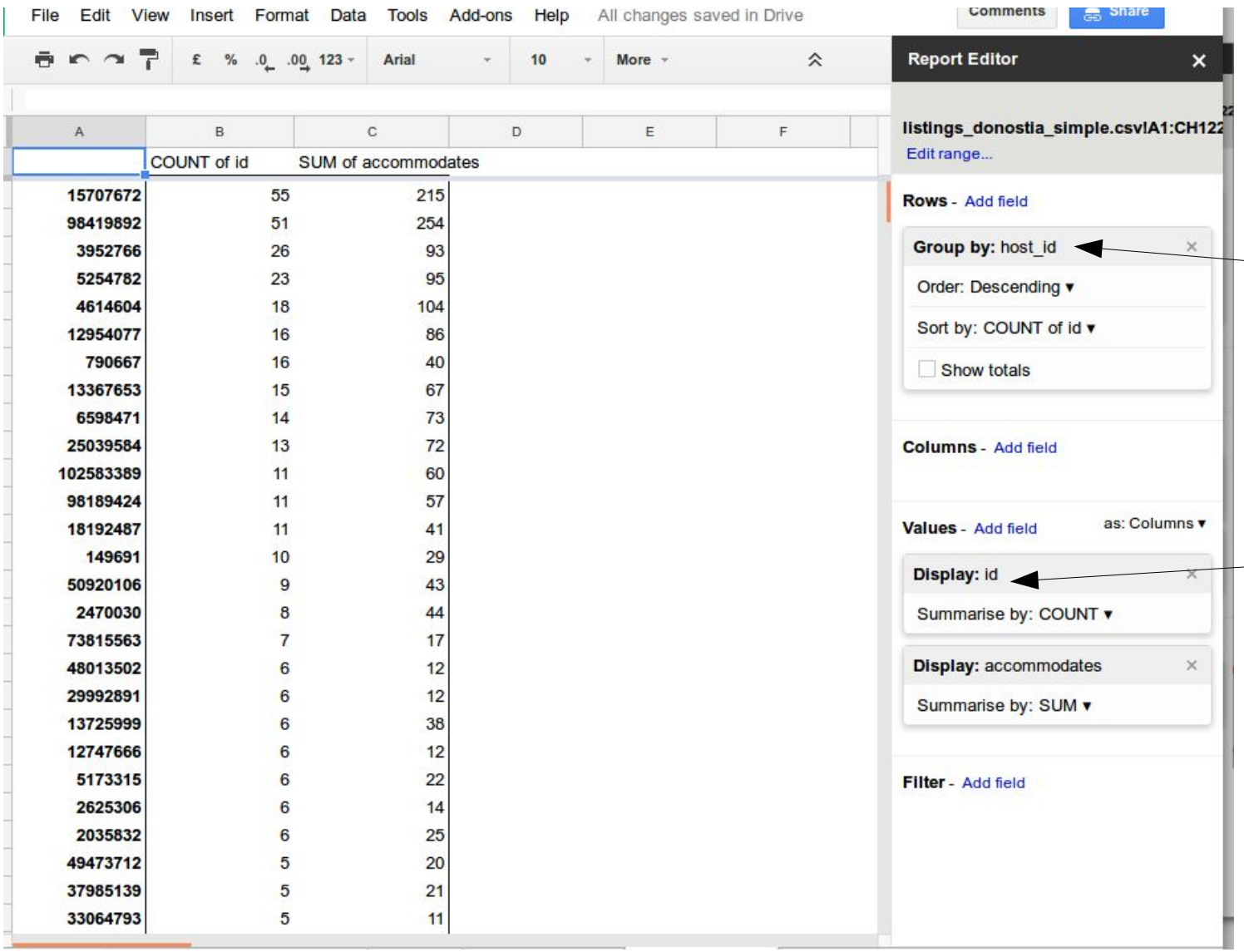
host_id

id (o cualquier variable) Summarise by COUNT



	A	B	C
1		COUNT of id	SUM of accommodates
2	98189424	11	57
3	18192487	11	41
4	3952766	8	18
5	13725999	6	38
6	76289581	4	9
7	50920106	3	14
8	39152861	3	6
9	22467824	3	8
10	6598471	3	26
11	115372864	2	8
12	106481544	2	4
13	92414536	2	4
14	36268260	2	4
15	35700062	2	4
16	8462790	2	5
17	6520043	2	7
18	4614604	2	10
19	2676178	2	4
20	120231094	1	6
21	115134693	1	3
22	115129584	1	6
23	109780975	1	4
24	109347536	1	3
25	106914989	1	5
26	106884120	1	12
27	105917109	1	5
28	105237788	1	4

Filtra por con y sin licencia (true/false)



	COUNT of id	SUM of accommodates
15707672	55	215
98419892	51	254
3952766	26	93
5254782	23	95
4614604	18	104
12954077	16	86
790667	16	40
13367653	15	67
6598471	14	73
25039584	13	72
102583389	11	60
98189424	11	57
18192487	11	41
149691	10	29
50920106	9	43
2470030	8	44
73815563	7	17
48013502	6	12
29992891	6	12
13725999	6	38
12747666	6	12
5173315	6	22
2625306	6	14
2035832	6	25
49473712	5	20
37985139	5	21
33064793	5	11