

Session 1

20/01/2025



GitHub repo

<https://github.com/CartoDB/research-public/tree/master/el-mundo-visualization-workshop>



CARTO setup

app.carto.com

Agenda:

17:00 - 17:45

Intro to CARTO & Spatial Data Visualization

17:45 - 19:00

Data Visualization Workshop: Analyzing Real Estate listings in Madrid with CARTO Builder

19:00 - 19:30

Break

19:30 - 21:00

Open Lab Time

Session 2

27/01/2025



GitHub repo

<https://github.com/CartoDB/research-public/tree/master/el-mundo-visualization-workshop>



CARTO setup

app.carto.com

Agenda:

17:00 - 17:30

CARTO Workflows Basics

17:30 - 19:00

Advanced Data Visualization Workshop:
Assessing the damages of La Palma Volcano

19:00 - 19:30

Break

19:30 - 21:00

Open Lab Time

Mastering Big Data Visualization & Storytelling

CARTO





Lucía García-Duarte

Data Scientist



Diego Vicente

Data Scientist



Miguel Álvarez

Data Science Lead

Agenda

Day 1

Intro to CARTO

Intro to Spatial Data Visualization

Data Visualization Workshop: Analyzing Real Estate Listings in Madrid

Open Lab

Agenda Day 2



CARTO Workflows basics

**Advanced Data Visualization Workshop:
Assessing the damages of La Palma Volcano**

Open Lab

Intro to CARTO

CARTO

Unlock the power of spatial analysis

1,200

Customers

170+

Team members

Accel

EARLYBIRD





The leading platform for
Location Intelligence and
Spatial Data Science



With an end to end platform:



Technology

Managed cloud or
on-premises
platform



Data

Open and
premium location
data streams

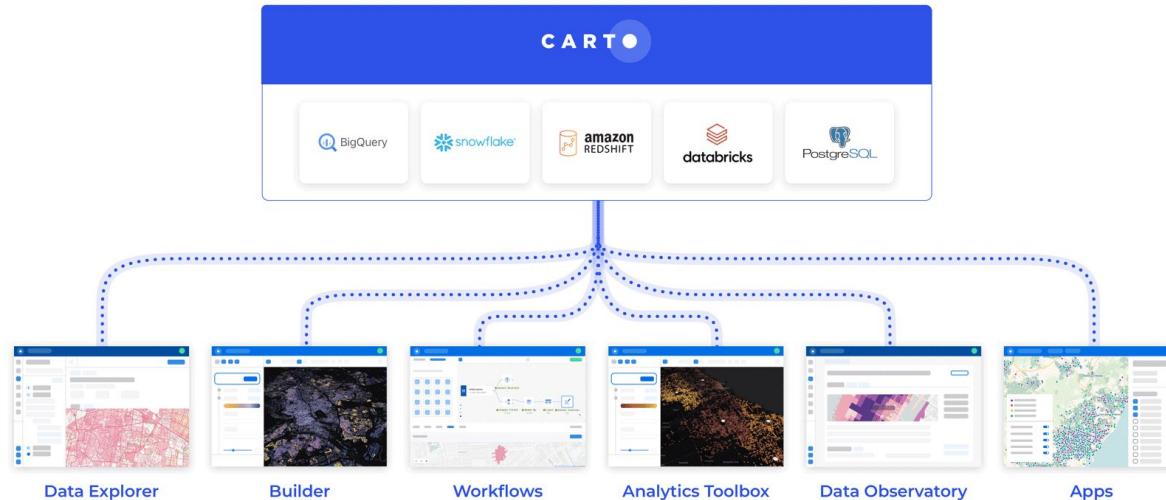


Services

Ongoing enablement
programs and custom
engagements

CARTO Cloud Native

CARTO brings together cloud connectivity, visualization, spatial analysis and development capabilities in a unified workspace to **unlock the power of spatial analysis**

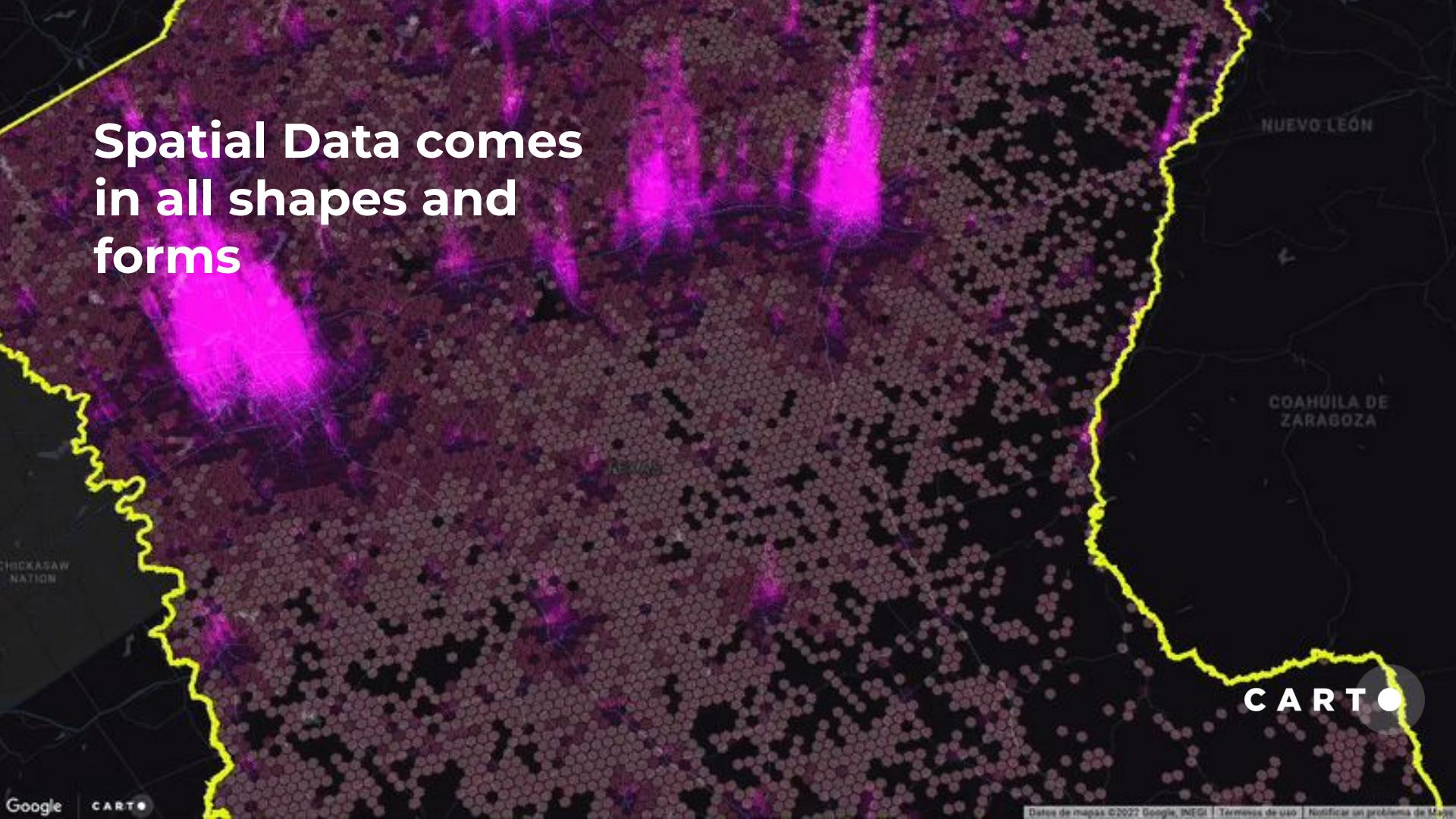


CARTO Platform. Guided Tour

<https://app.carto.com/>

The screenshot displays the main interface of the CARTO Platform. At the top left is the 'Workspace' sidebar with links to Home, Maps, Workflows, Data Explorer, Data Observatory, Connections, Applications, Developers, and Settings. The main content area features a 'Welcome to CARTO' message encouraging users to start creating maps and performing spatial analysis. Below this are three examples of spatial analysis: 'Pinpoint new store locations closest to your customers' (3 weeks ago), 'Monitor retail store performance' (8 min. ago), and 'Analyzing Airbnb ratings in Los Angeles' (8 min. ago). Further down, there's a section titled 'Start creating workflows' showing various workflow templates such as 'Invite users', 'Calculate distance between two points', 'Get all stores near a location', and 'Get all stores near a location'. A 'Create map' button is also visible.

Understanding Spatial Data: Key definitions and Concepts



**Spatial Data comes
in all shapes and
forms**

CARTO

Spatial Data classification

1. Data aggregated by administrative boundaries

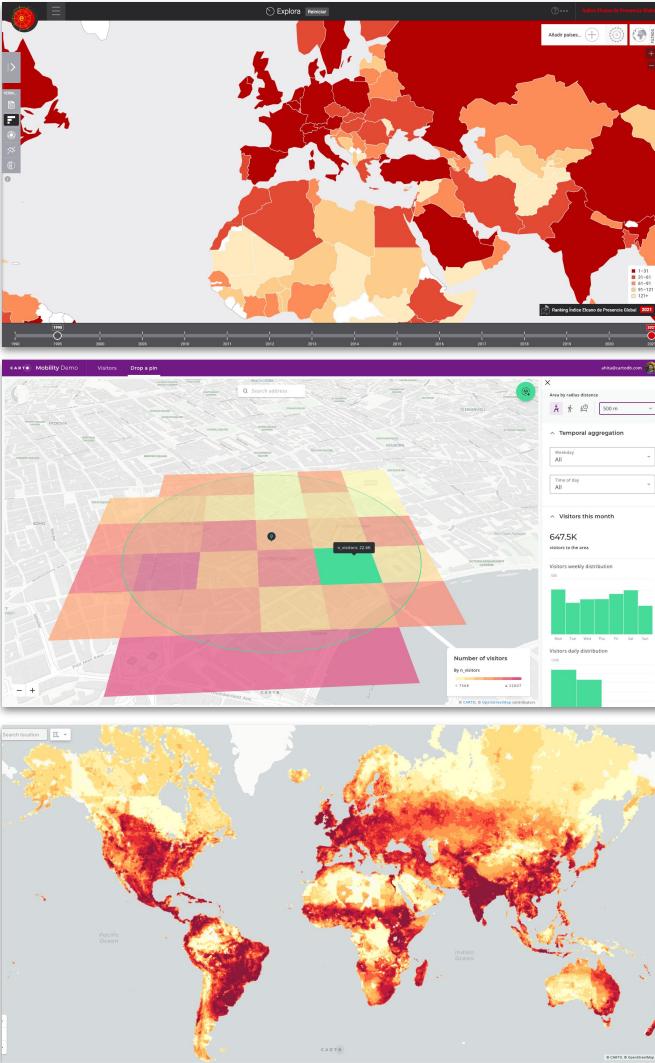
Is used to monitor and assess the status of and changes in administrative territories.

2. Data derived from a Position (pair of coordinates)

Human activity data is used to monitor and assess the status of and changes in human behaviour.

3. Data derived from Earth Observation satellites

Earth observation is used to monitor and assess the status of and changes in natural and built environments.



Why is Spatial special? Spatial Dependence

"Everything is related to everything else, but near things are more related than distant things."

(Tobler, 1970)

CARTO — Unlock the power of spatial analysis.

Introduction to Data Visualization



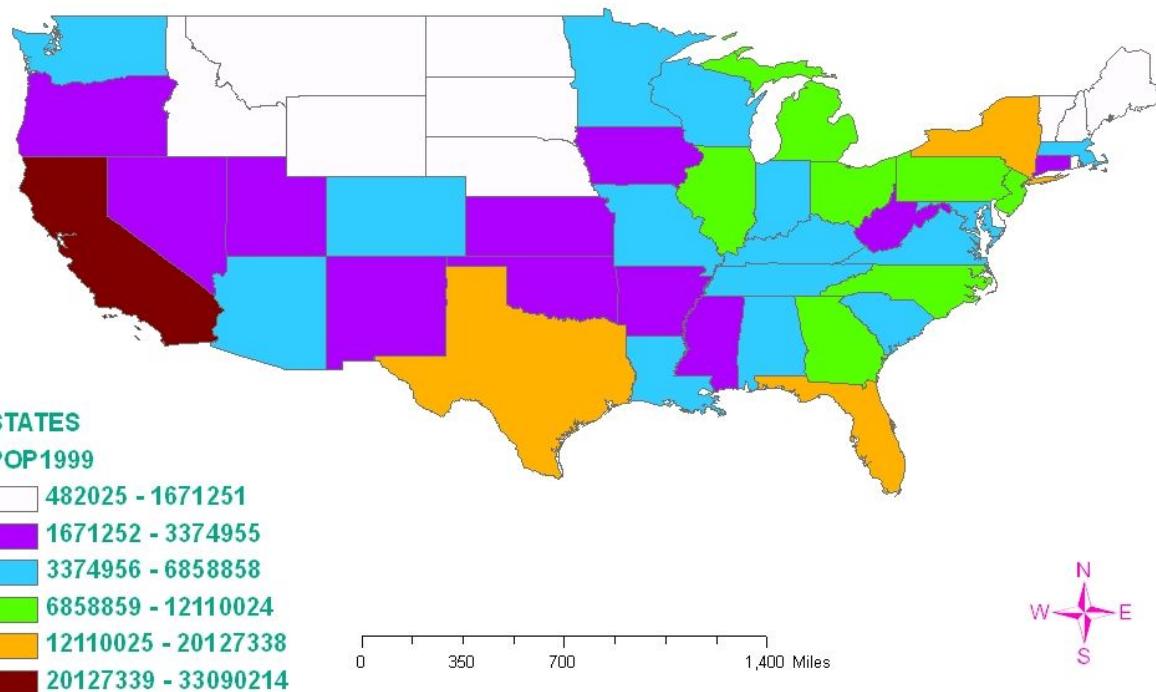
**Great dataviz isn't always
about making beautiful
maps**

(that's a bonus!)

**It's about answering questions
clearly and effectively**

**Remember,
just because
you
understand
your map,
doesn't mean
everyone
will...**

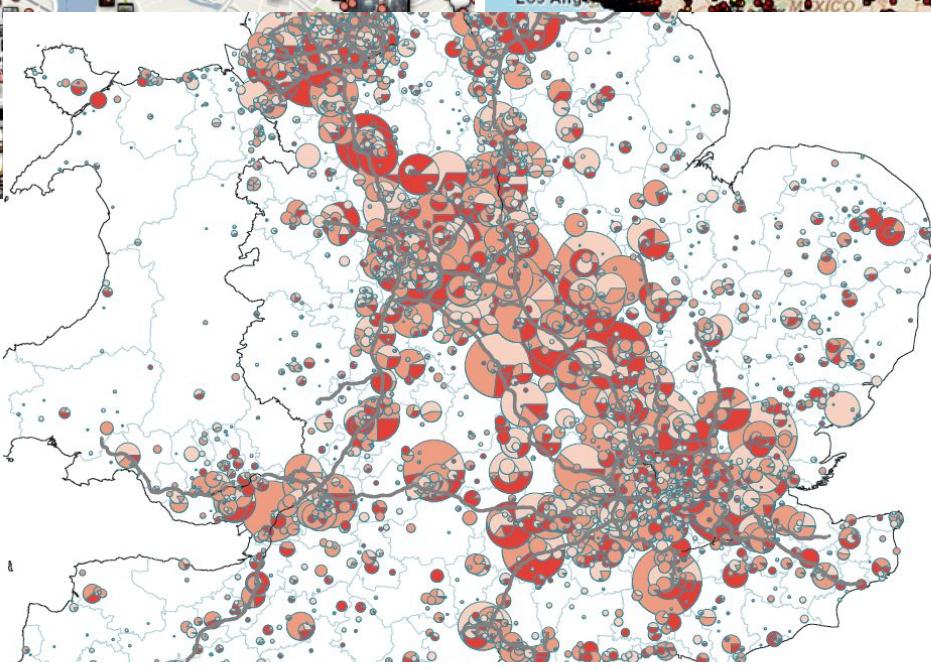
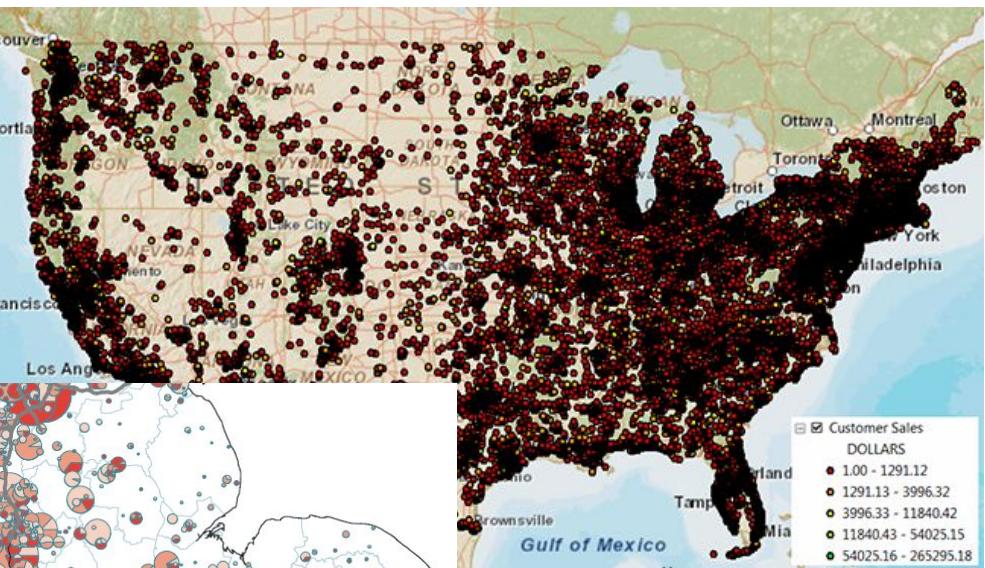
Population Trend Bad Harmony Map





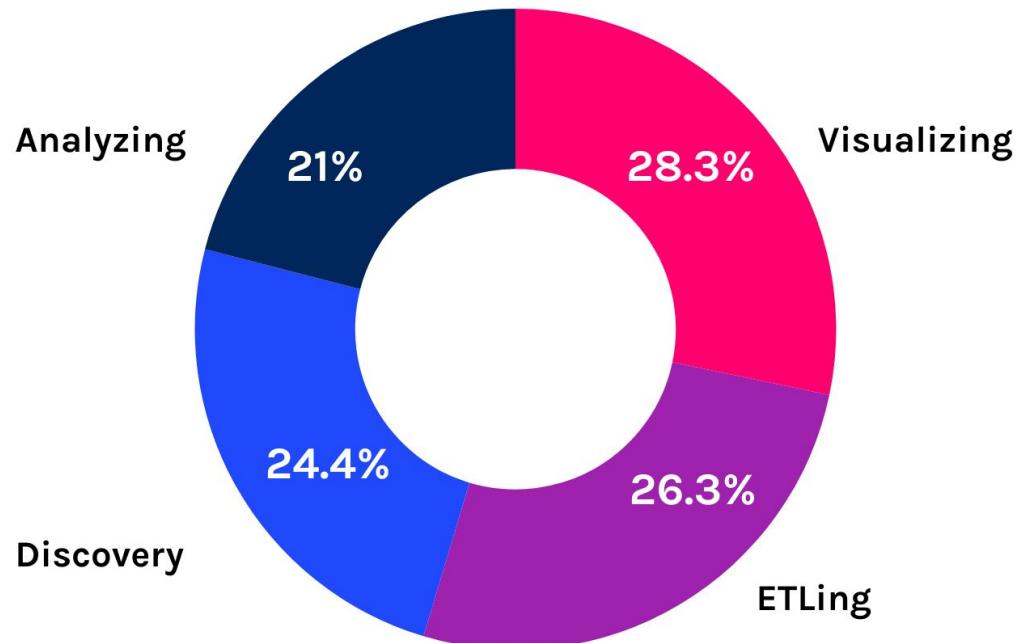
You don't notice
good design.





The State of Spatial Data Science
Survey 2024

What is the most difficult part of working with spatial data?



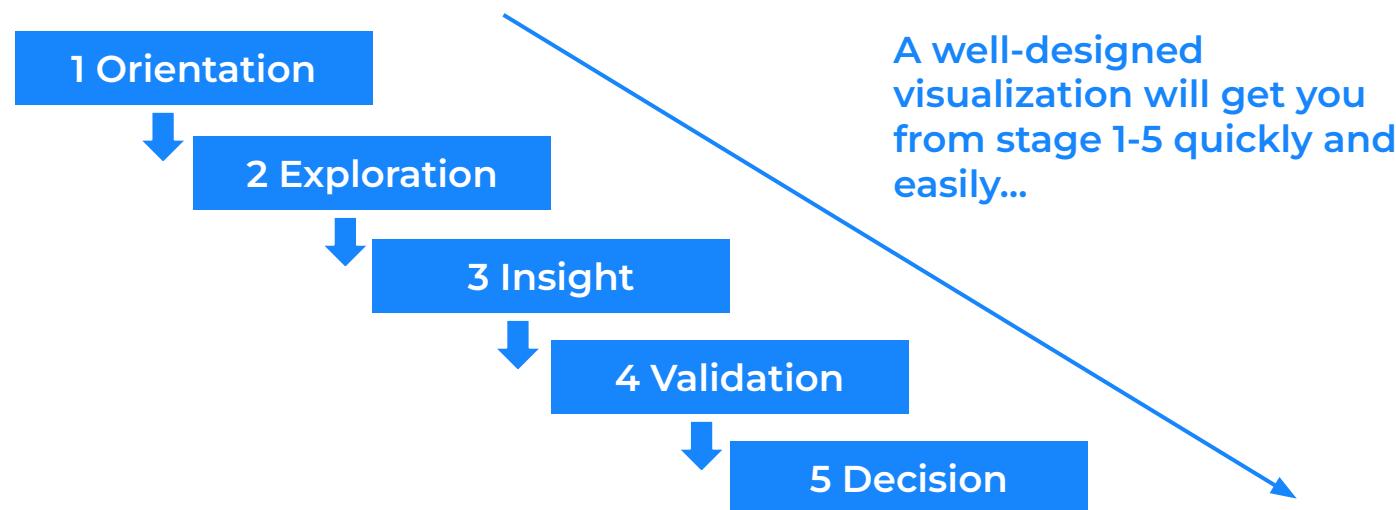
Why is visualizing data so difficult?

1. Priority
2. Subjectivity
3. Technical constraints
4. Data size and complexity
5. Overloading the map
6. Balancing beauty with user needs

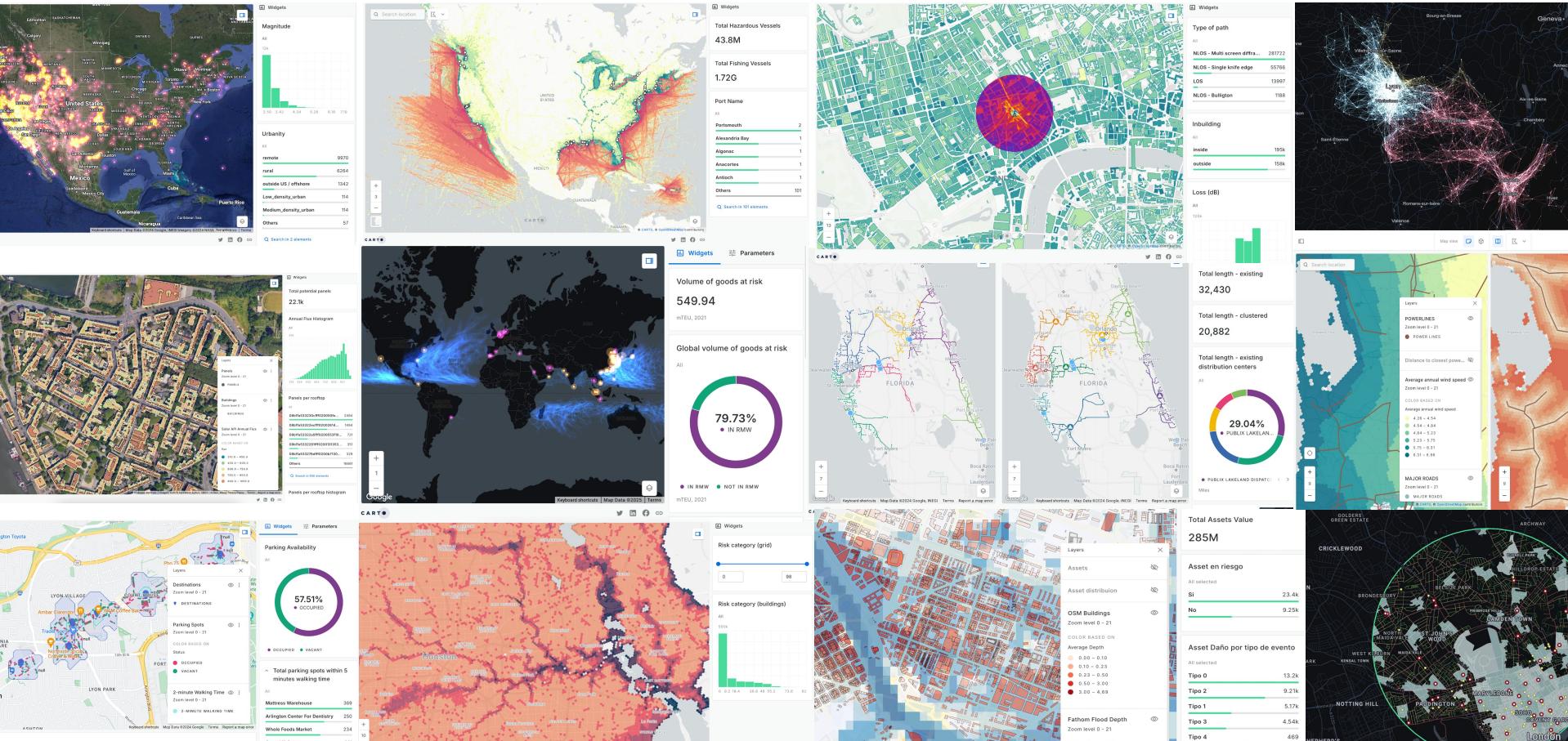


The data visualization user journey

Let's think about what a user does when they open a map...



Some CARTO inspo

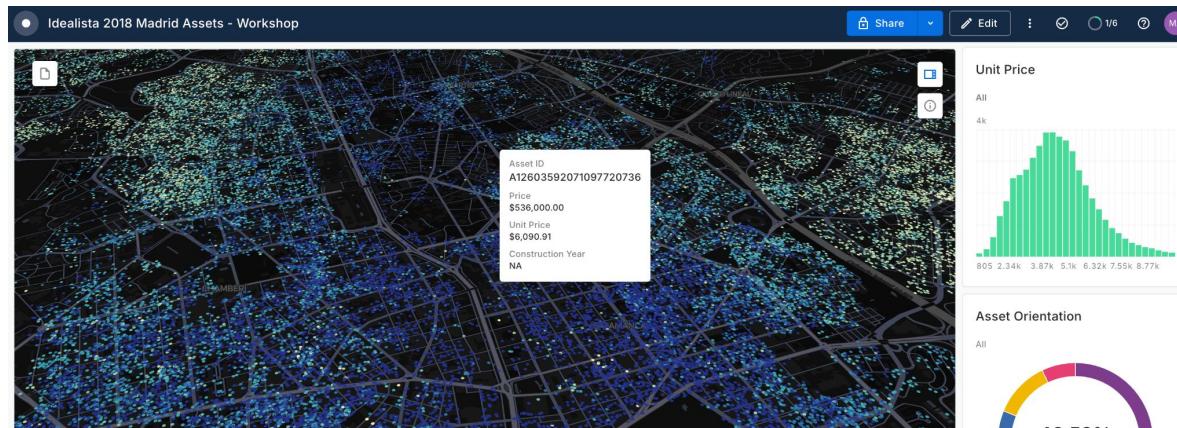


Data Visualization Workshop: Analyzing Real Estate listings in Madrid with CARTO Builder

[Access full guide:](#)

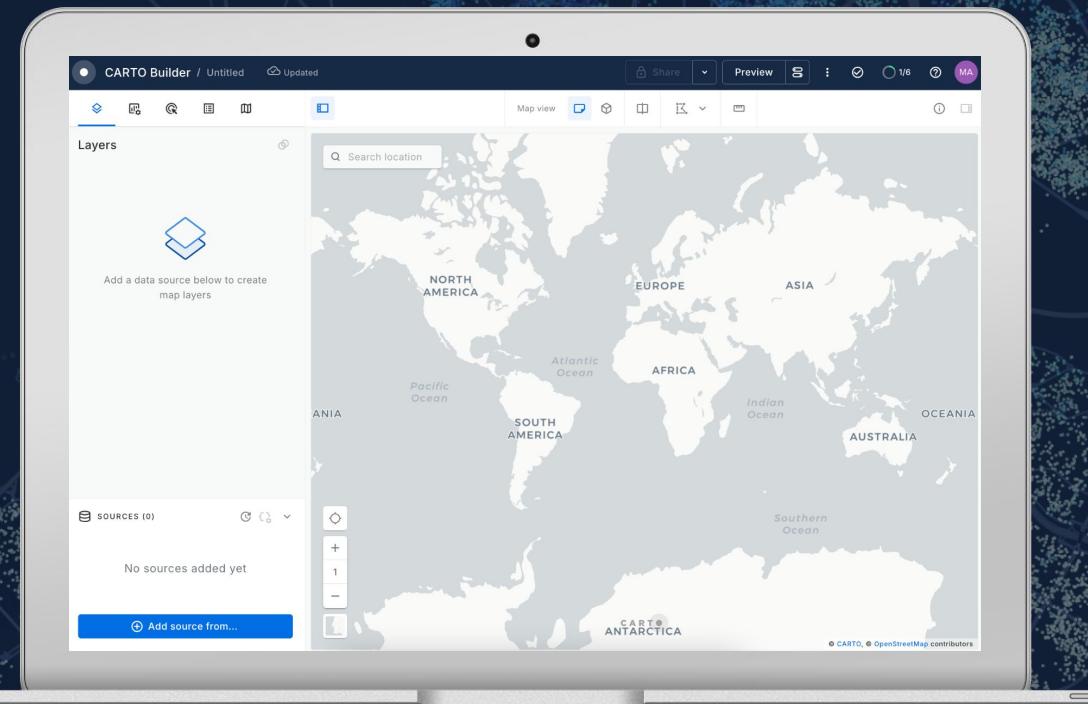
Idealista18 dataset

- This [dataset](#) contains information about 2018 real estate listings in Spain
 - + **90,000 real estate listings** in the city of Madrid
- Enriched with **geographic** and **property-specific** data
 - property prices
 - indoor characteristics (e.g., size, number of rooms),
 - geographical features such as proximity to urban points of interest



CARTO — Unlock the power of spatial analysis

Open Lab Time

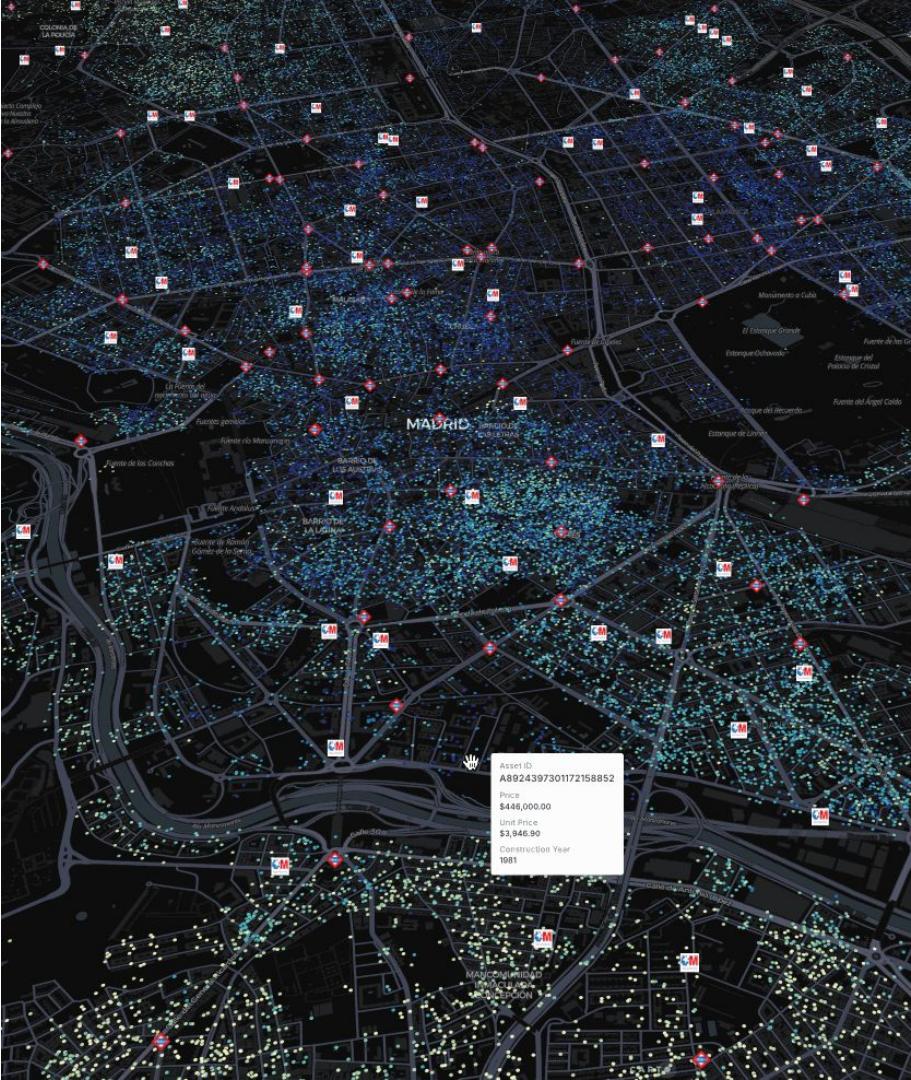


Modify & improve our Real Estate Listings map...

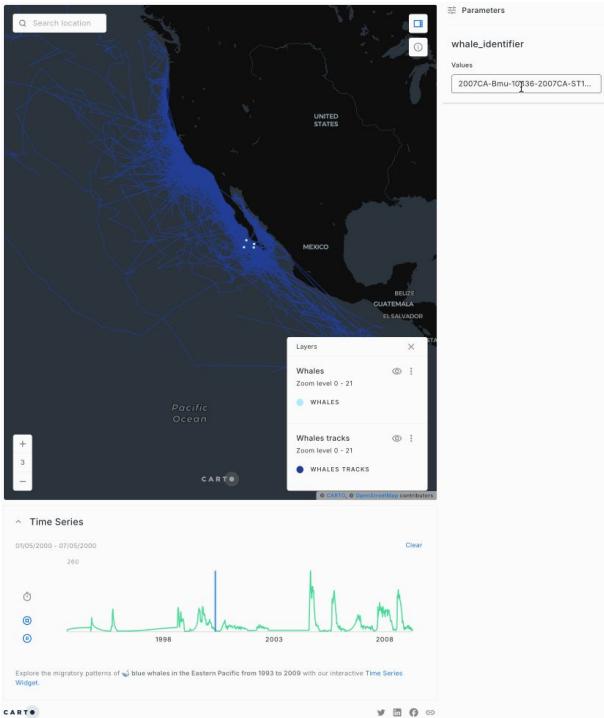
- Change the styling: try new palettes, change the radius of the points, add labels to the neighbors...
- Add new widgets to extract more insights
- Make your map more appealing by including more layers:
 - *Metro Stations* (from the [Portal de Datos Abiertos Consorcio Regional de Transportes de Madrid](#)). Download the **shapefile** [here](#) and upload it to CARTO
 - *Health Care Centers* (from the [Portal de datos abiertos del Ayuntamiento de Madrid](#)). Download the **csv file** [here](#) and upload it to CARTO



You all belong to the same organization and are using the same CARTO DW. Be **careful when uploading data** to avoid overwriting your colleagues tables!



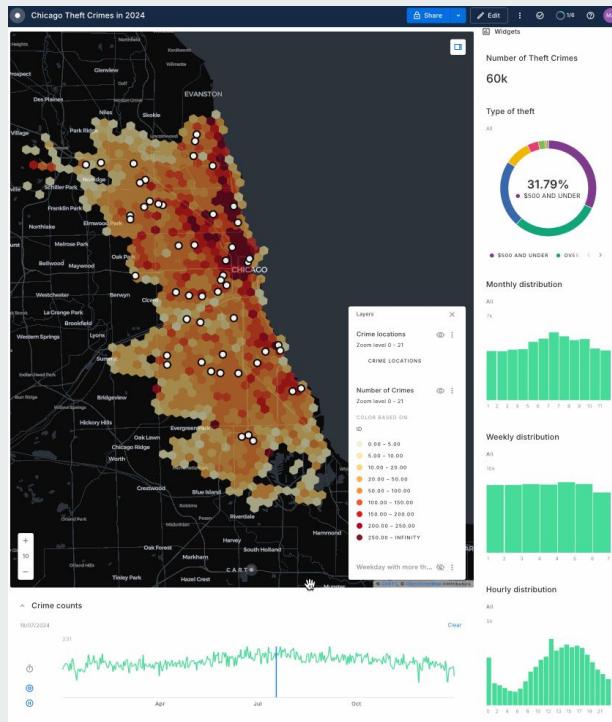
The Migration of Blue Whales in Eastern Pacific



```
@ CARTODW > carto-demo-data > demo_tables
> blue_whales_eastern_pacific_point
> blue_whales_eastern_pacific_line
```

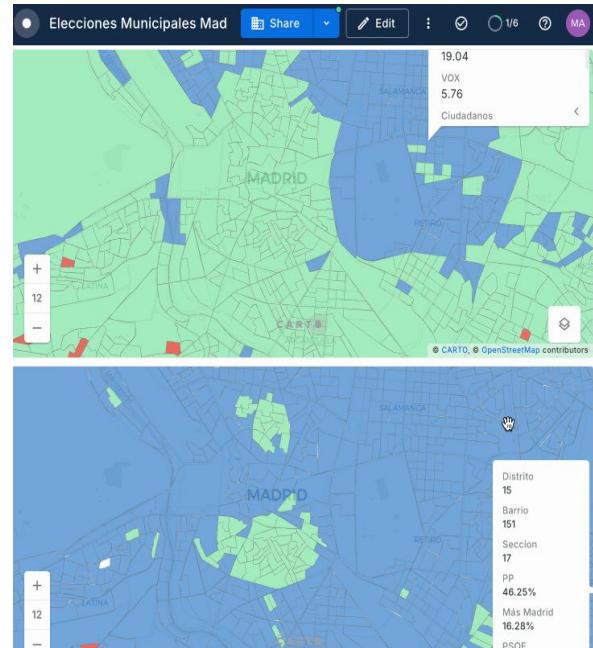
Or create a map from scratch...

Chicago Theft Crimes in 2024



```
@ CARTODW > organization data > shared_us
> chicago_thefts_2024
```

Comparison of Madrid's Election Results



```
@ CARTODW > organization data > shared_us
> elecciones_municipales_2019
> elecciones_municipales_2023
@ cartobq.docs.barrios_madrid
```

Story-telling with Low-code Spatial Analytics

CARTO Workflows



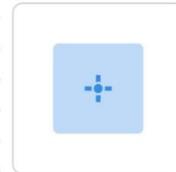
H3
center



Point Stats
In Polygons



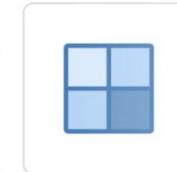
Create
Tileset



Quadbin
center



Quadbin
Polyfill



Quadbin
center



H3
Center



Point SI
In Polyg



Nearest
Neighbors



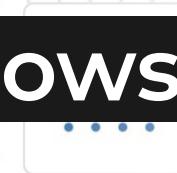
ST Cluster
K-Means



ST
Voronoi



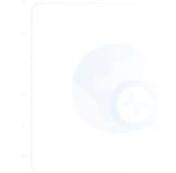
What is Workflows,
exactly?



ST Cluster
DBSCAN



ST
Difference



K-Near
Neighb



Statistics



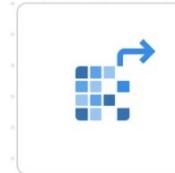
H3 from
GeoPoint



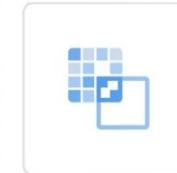
H3
Boundary



H3
Polyfill



Get values
from Raster



Intersect &
aggregate



Statistics



H3 fro
GeoPo

Workflows turns this:

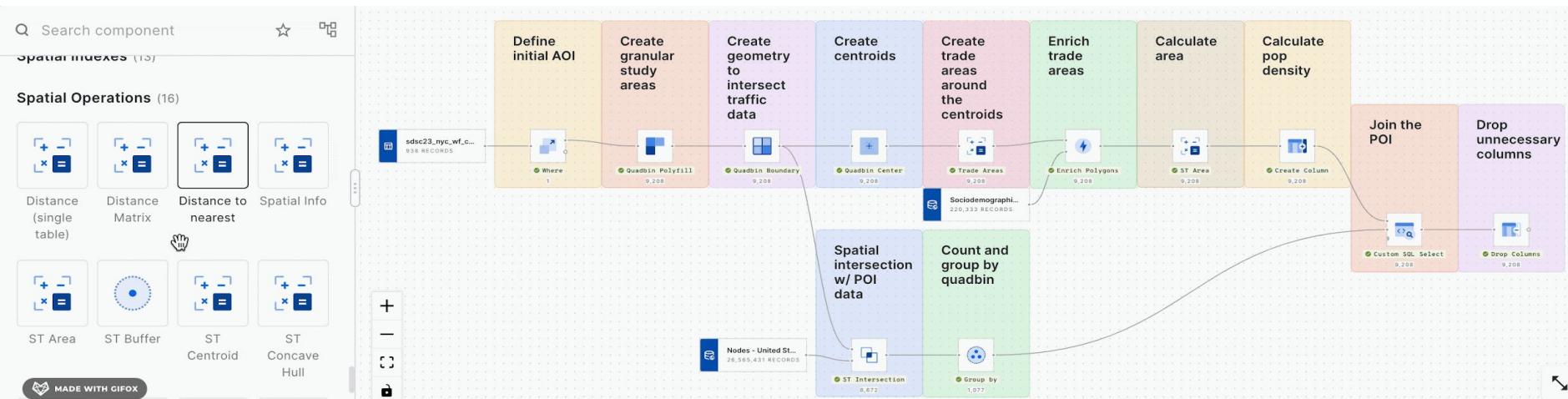
```

328
329 /*===== native.dropcolumn (v1) [a75a1997-0a5d-4d5f-bd57-7ef7bf40e83a] =====*/
330 BEGIN
331 CREATE TABLE IF NOT EXISTS `cartodb-gcp-solutions-eng-team.workflows_temp.WORKFLOW_aca13cc597bbcd3e_54119db088c34f5b_result`
332 OPTIONS (
333   expiration_timestamp = TIMESTAMP_ADD(
334     CURRENT_TIMESTAMP(), INTERVAL 30 DAY
335   )
336 )
337 AS
338   SELECT * EXCEPT (geom_trade_area, radius_size, quadbin_geo, geom)
339   FROM `cartodb-gcp-solutions-eng-team.workflows_temp.WORKFLOW_aca13cc597bbcd3e_aea4e488231654e6_result`;

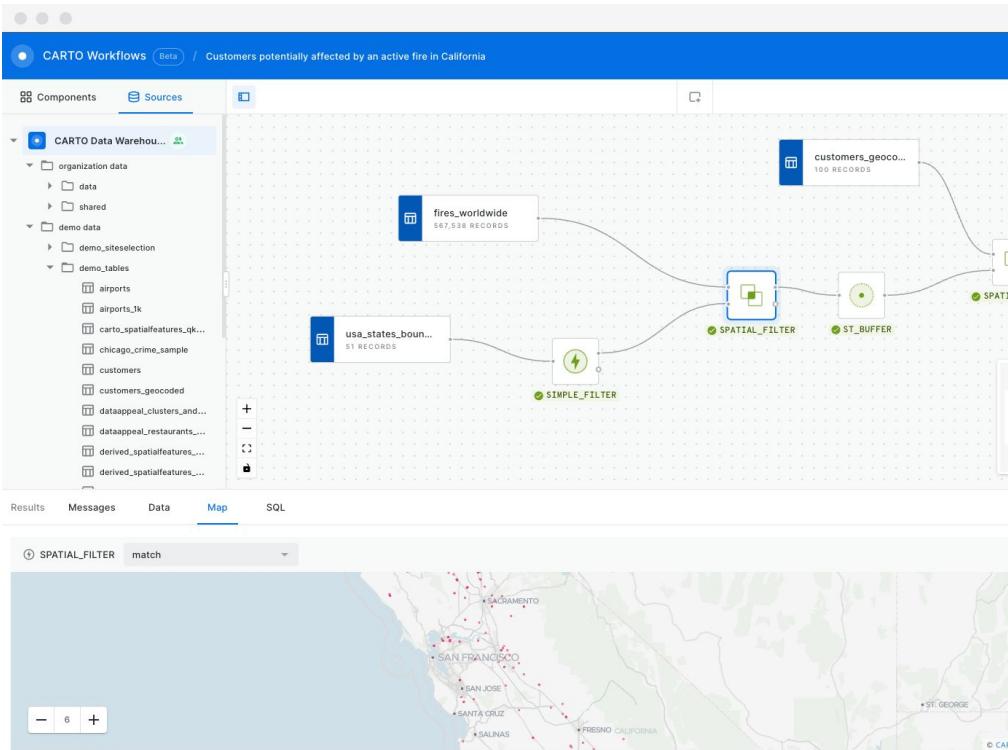
```

MADE WITH GIFOX

Into this:



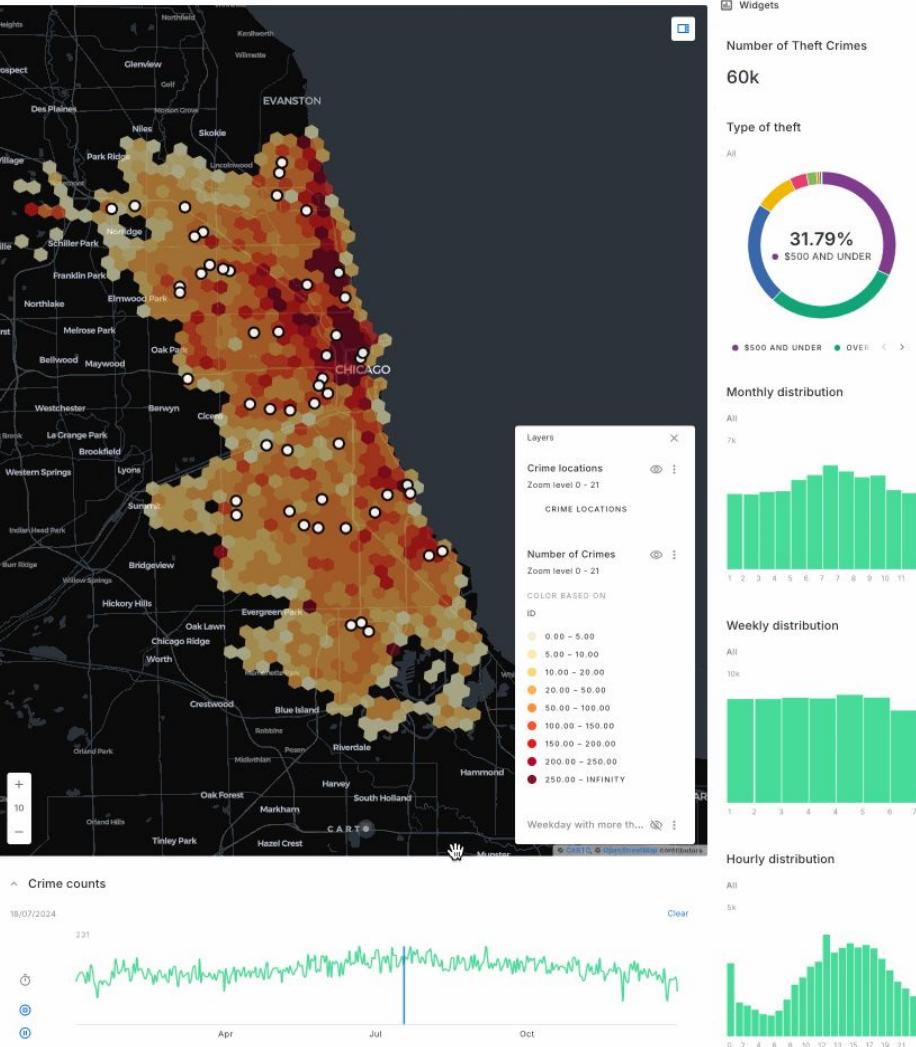
- Advanced functionality for skilled users to create pipelines
- Easy to collaborate regardless of technical level
- Suite of tools to help preview, debug, and automate an analysis
- Configurable components for a full range of analytical functions
- Reduce the time to learn new environment-specific nuances



Remember this map?

The data source was first processed using
Workflows:

- Missing dates & coordinates
 - No `geom` column
 - ...



Data

Chicago Crimes

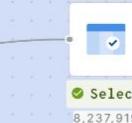
Load the [Chicago Crime](#) BQ's public table, which reflects reported incidents of crime that occurred in the City of Chicago from 2001 to present, minus the most recent seven days. Data is extracted from the Chicago Police Department's CLEAR (Citizen Law Enforcement Analysis and Reporting) system.



Cleaning & Filtering

Process and filter the data, and store the results in a table to create a map

Select variables of interest



Remove crimes with missing date or location



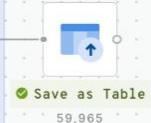
Extract geometries from lat/lon coordinates



Filter by theft crimes committed in 2024



Get hour, weekday and month of crime from date



Workflows are made of components...

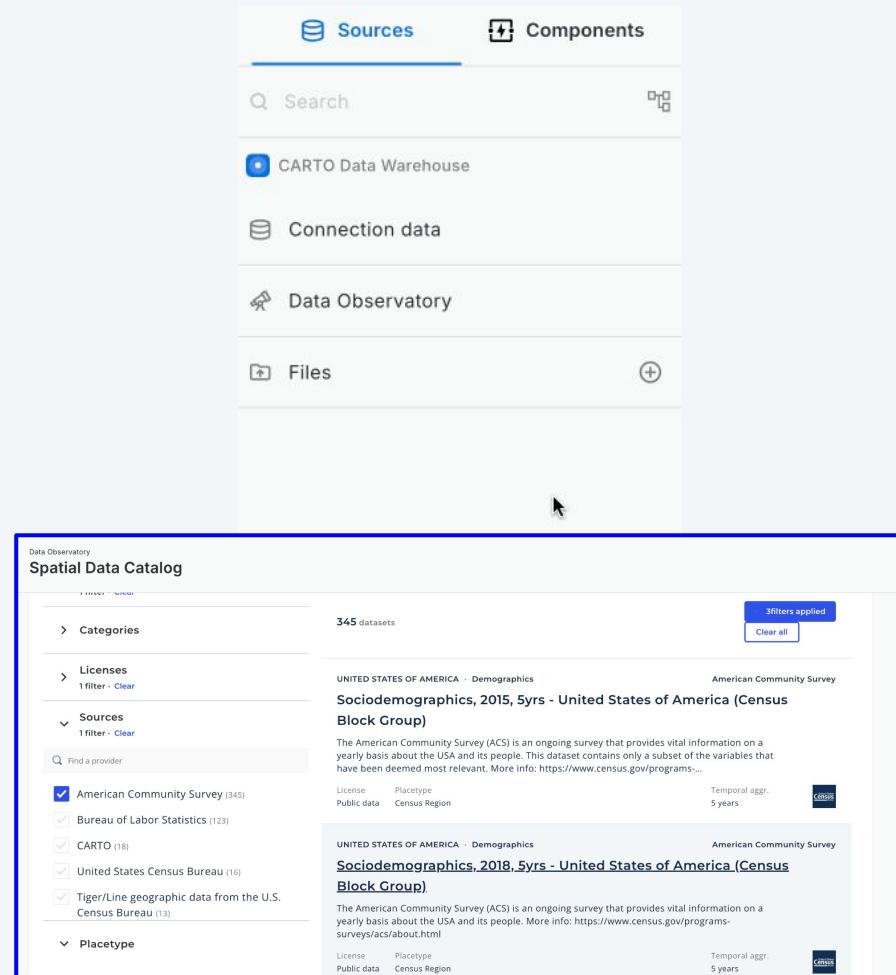
- Aggregation (6)
- Custom SQL (2)
- Data Enrichment (4)
- Data Preparation (30)
- Generative AI (1)
- **Input/Output (7)**
- Joins (8)
- Geo Parsers (6)
- Raster Operations (2)
- Spatial Accessors (5)
- Spatial Analysis (6)
- Spatial Constructors (16)
- **Spatial Indexes (13)**
- Spatial Operations (17)
- Spatial Statistics (11)
- **Tileset Creation (4)**

More components added regularly! | <https://docs.carto.com/whats-new>

...and Sources

Data you bring to your workflow for analysis activities.

They can be from your warehouse, from the [Spatial Data Catalog](#), an external link, a csv, or created by a component.



The screenshot shows the CARTO Data Observatory interface. At the top, there are tabs for "Sources" (selected) and "Components". Below the tabs is a search bar and a "Sources" icon. The main area displays a list of data sources:

- CARTO Data Warehouse** (with a blue icon)
- Connection data** (with a database icon)
- Data Observatory** (with a globe icon)
- Files** (with a folder icon)

On the right side of the interface, there is a "+" button. A mouse cursor is visible near the top right corner. Below the main interface, a modal window titled "Spatial Data Catalog" is open, also showing the "Sources" tab. This modal lists datasets and filters applied:

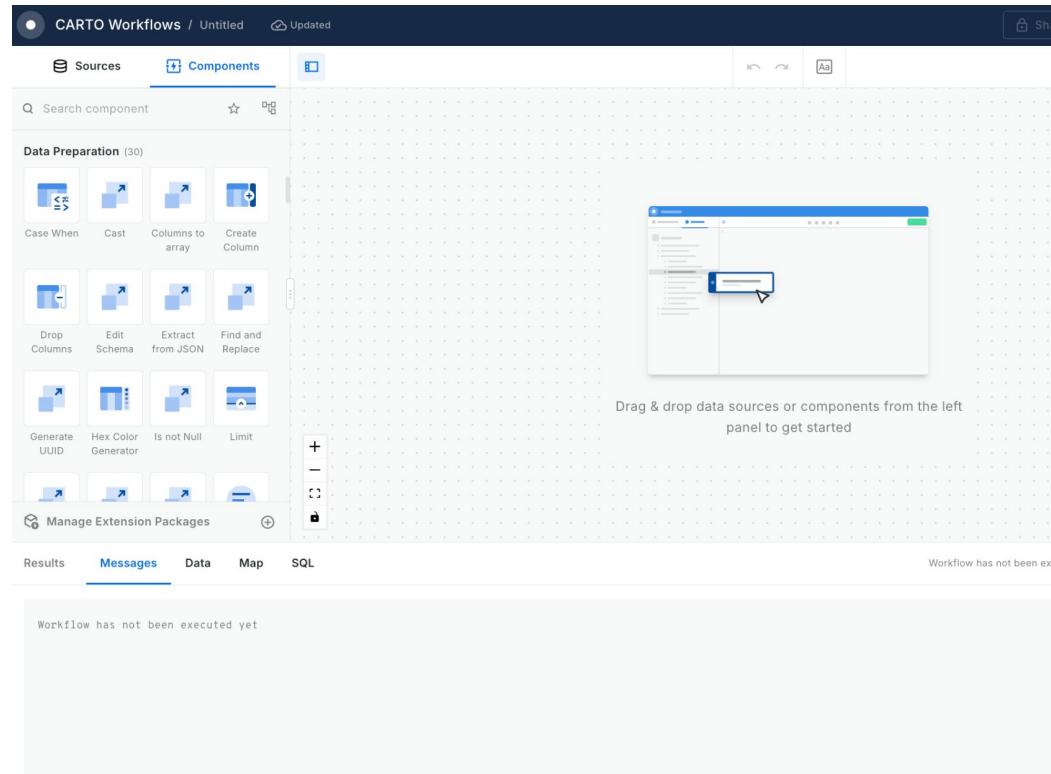
- 345 datasets**
- Filters applied** (button)
- Clear all** (button)

The modal content includes:

- Categories**: American Community Survey, Bureau of Labor Statistics, CARTO, United States Census Bureau, Tiger/Line geographic data from the U.S. Census Bureau.
- Licenses**: 1 filter - Clear
- Sources**: 1 filter - Clear
- Find a provider** search bar
- American Community Survey** dataset details:
 - UNITED STATES OF AMERICA - Demographics
 - Sociodemographics, 2015, Syrs - United States of America (Census Block Group)**
 - Description: The American Community Survey (ACS) is an ongoing survey that provides vital information on a yearly basis about the USA and its people. This dataset contains only a subset of the variables that have been deemed most relevant. More info: <https://www.census.gov/programs-surveys/acs/about.html>
 - License: Public data
 - Placetype: Census Region
 - Temporal aggr.: 5 years
 - Census** logo
- American Community Survey** dataset details:
 - UNITED STATES OF AMERICA - Demographics
 - Sociodemographics, 2018, Syrs - United States of America (Census Block Group)**
 - Description: The American Community Survey (ACS) is an ongoing survey that provides vital information on a yearly basis about the USA and its people. More info: <https://www.census.gov/programs-surveys/acs/about.html>
 - License: Public data
 - Placetype: Census Region
 - Temporal aggr.: 5 years
 - Census** logo

A quick look

[https://clausa.app.carto.com/
workflows](https://clausa.app.carto.com/workflows)



Importing data from a Google Drive spreadsheet

To import a specific Google Drive sheet, use this URL:

```
https://docs.google.com/spreadsheets/d/{key}/gviz/tq?tqx=out:csv&sheet={sheet_name}
```

- `key` is your document's ID
- `sheet_name` is the name of the sheet you want to import into Workflows

Example:

Google Drive spreadsheet link:

```
https://docs.google.com/spreadsheets/d/abc1234567/edit#gid=0
```



URL to import data into Workflows:

```
https://docs.google.com/spreadsheets/d/abc1234567/gviz/tq?tqx=out:csv&sheet=Sheet1
```

More advanced tools

Visualizing large datasets using Tilesets

Tileset Creation (4)



Create H3
Agg Tileset



Create
Point Agg
Tileset

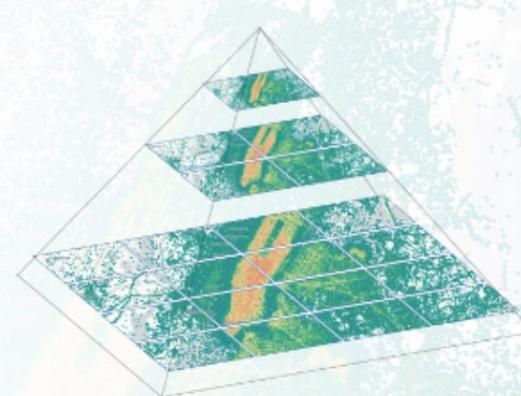


Create
Quadbin
Agg Tileset

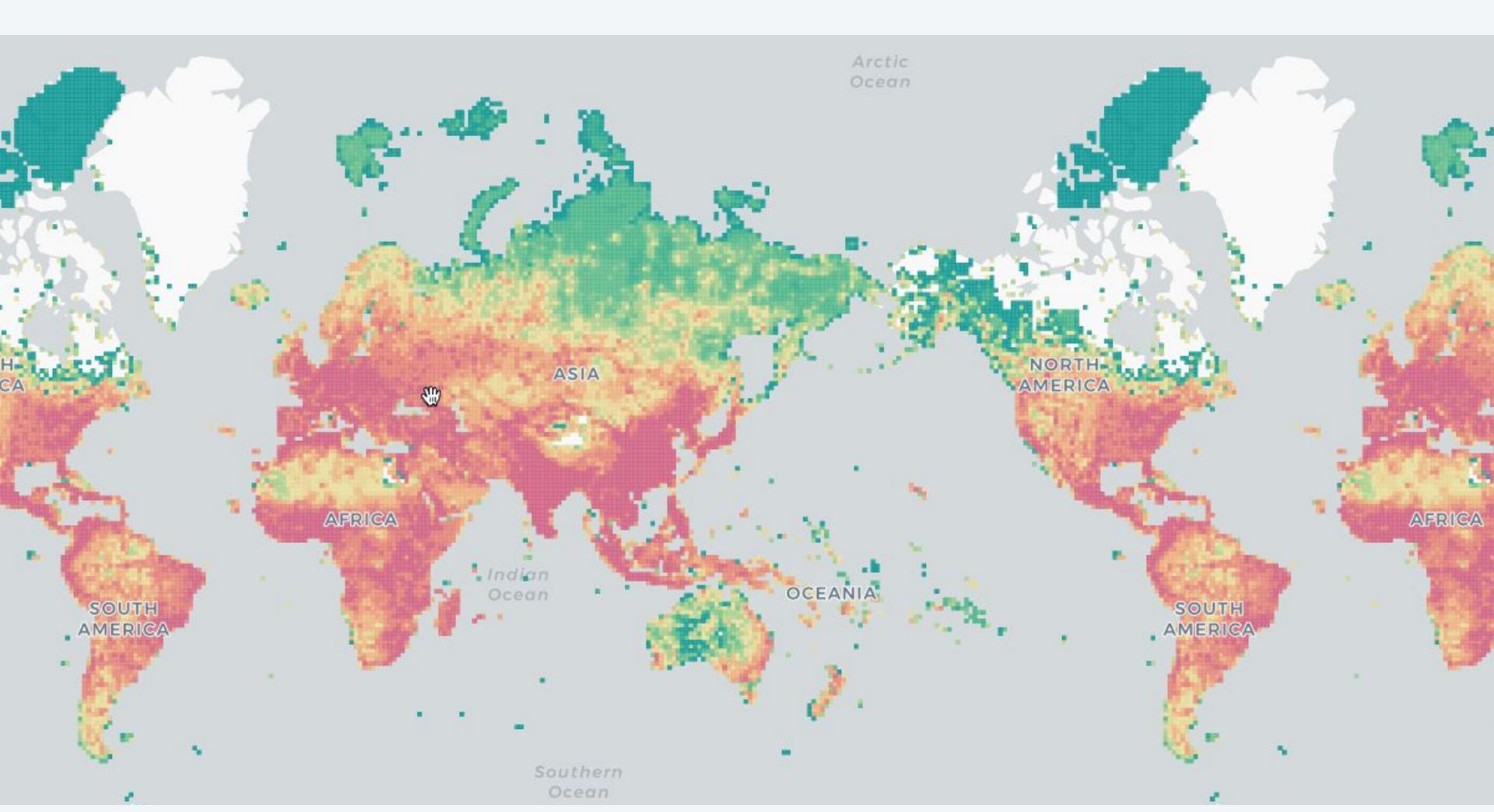


Create
Vector
Tileset

A vector tileset is a representation of the data optimized for map visualizations, which allows to process and visualize massive spatial datasets with CARTO natively in your data warehouse.



Tilesets are generated in SQL using the procedures available in the tiler module of the [Analytics Toolbox](#) for the different data warehouse platforms. The type of tileset will depend on the nature of the geometries in your source table.



SQL Parameters

Create a SQL Parameter

SQL Parameters are placeholders that can be used on any SQL Query data source in Builder.

After creating a new SQL Parameter, it needs to be added manually to one or more SQL Query data sources. Once added, the actual value for the parameter can be defined through a control UI in the right side panel's 'Parameters' tab.

Learn more about SQL Parameters in our [Documentation](#).

```
1  SELECT
2    carto.H3_FROMGEOPODTH3(h3, h3,
3      AVG(value) AS average_value
4    ) AS average_value
5    FROM events
6    WHERE event_time > {{date_from}}
7      AND event_time < {{date_to}}
8    GROUP BY h3
9    WHERE type IN {{type_category}}
10   GROUP BY 1
11   ORDER BY average_value DESC;
```

Create a SQL Parameter

Select the type of parameter:



Date

Pick start and end dates from a calendar that can be used to filter by a date range in your query.

[More info](#)



Numeric

Use a simple or range slider to retrieve a unique or a pair of numeric values to be used in your query.

[More info](#)



Text

Type or select values from a predefined list that will be interpreted as strings in your query.

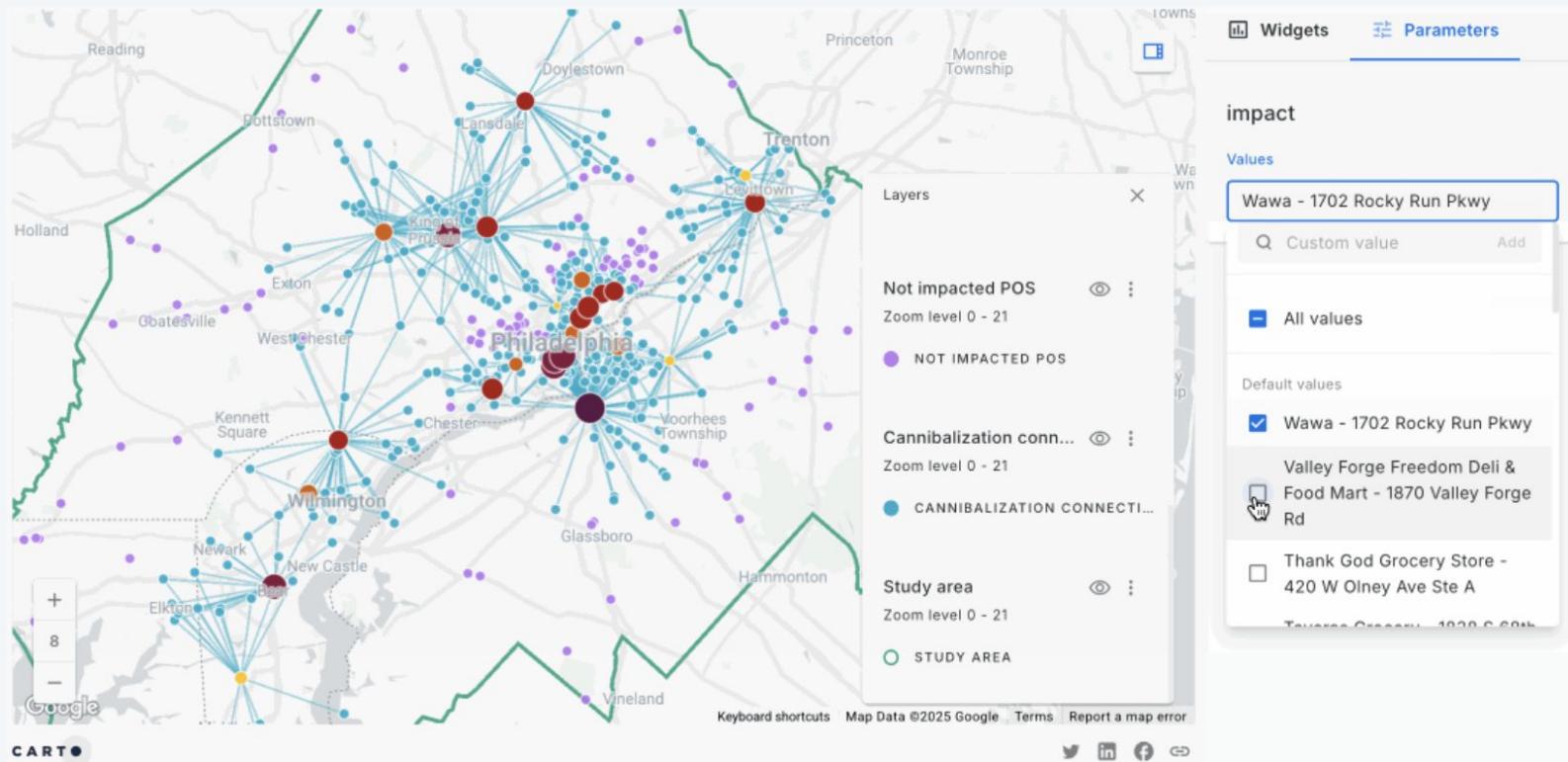
[More info](#)

Don't show again

Continue

Back

Continue



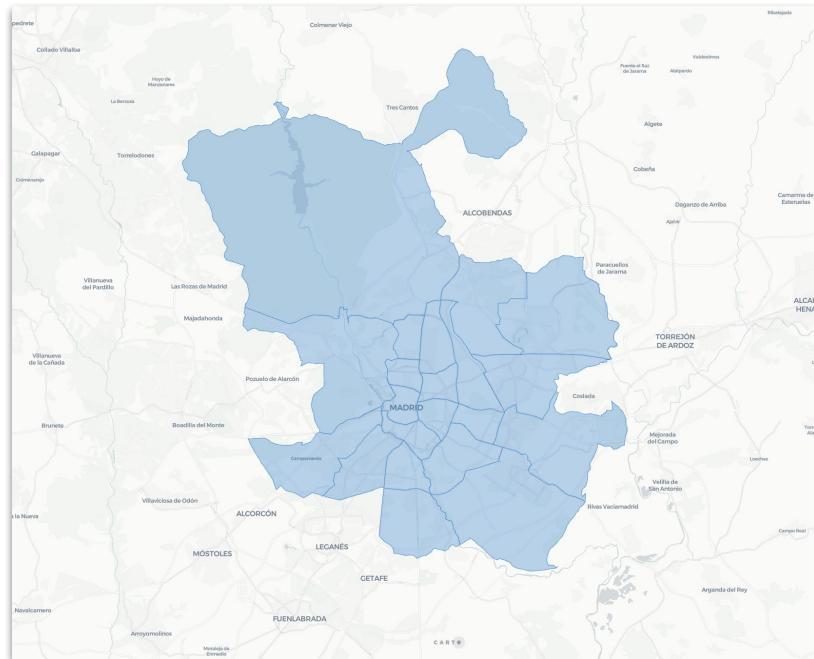
Basic SQL Review



```
SELECT
  * EXCEPT (geom)
FROM
  cartobq.docs.madrid_districts
```

id	name	area
9	Moncloa - Aravaca	46531136.1514
1	Centro	5228246.7414
6	Tetuán	5374724.44595
7	Chamberí	4679185.04678
17	Villaverde	20187586.3406
16	Hortaleza	27626096.0403
10	Latina	25427202.5964
5	Chamartín	9175481.878
13	Puente de Vallecas	14968606.1543
3	Retiro	5466211.16436

Starting with a basic SELECT





```
SELECT
  name,
  area,
  name IN (
    'Centro', 'Arganzuela',
    'Retiro', 'Salamanca',
    'Chamartín', 'Tetuán',
    'Chamberí'
  ) AS within_m30 -- (*)
FROM
  cartobq.docs.madrid_districts
```

name	area	within_m30
Moncloa - Aravaca	46531136.1514	false
Centro	5228246.7414	true
Tetuán	5374724.44595	true
Chamberí	4679185.04678	true
Villaverde	20187586.3406	false
Hortaleza	27626096.0403	false
Latina	25427202.5964	false
Chamartín	9175481.878	true
Puente de Vallecas	14968606.1543	false
Retiro	5466211.16436	true

Adding new features

- New features are added using the syntax: <expression> AS <alias>
- An expression includes:
 - Arithmetic
 - String operations
 - Conditions
 - Presence in collections
 - Function calls



WITH

```
m30_cte AS (      -- (1)
    SELECT
        *,
        name IN (
            'Centro', 'Arganzuela',
            'Retiro', 'Salamanca',
            'Chamartín', 'Tetuán',
            'Chamberí'
        ) AS within_m30
    FROM
        cartobq.docs.madrid_districts
)
SELECT
    name,
    area,
    within_m30
FROM
    m30_cte          -- (2)
```

Using Common Table Expressions

- A Common Table Expression or CTE lets you **name subqueries** (1) to **use them later** in the query (2).
- It usually helps with **readability**, but also lets you use variables that were not within scope before.
- They are **not cached** in BigQuery.



```
WITH
m30_cte AS (
    -- [...]
)
SELECT
    within_m30,          -- \
    COUNT(*) AS n_districts,   -- |(2)
    SUM(area) AS total_area   -- /
FROM
    m30_cte
GROUP BY
    within_m30           -- (1)
ORDER BY
    total_area DESC        -- (3)
```

within_m30	n_districts	total_area
false	14	562676698.3054
true	7	41778428.3339

Basics of grouping

1. GROUP BY syntax will **group all rows sharing the same values** in the provided key(s).
2. All variables mentioned in the SELECT statement must be **grouping keys or aggregations**.
3. **ORDER BY happens after GROUP BY**, so you can reference aggregations.



```
WITH
m30_cte AS (
    -- [...]
)
SELECT
    within_m30,
    COUNT(*) AS n_districts,
    SUM(area) AS total_area
FROM
    m30_cte
WHERE                                -- (1)
    name LIKE 'C%'
GROUP BY
    within_m30
HAVING                                -- (2)
    n_districts > 2
```

within_m30	n_districts	total_area
true	3	19082913.66618

Basics of filtering

1. To **filter rows present in the FROM source**, use the **WHERE** syntax.
2. To filter rows **after a GROUP BY aggregation**, use the **HAVING** syntax.

There is yet another syntax for filtering! (in the next slide)



```
WITH
  m30_cte AS (
    -- [...]
  )
SELECT
  name,
  area,
  within_m30,
  RANK() OVER (          -- (1)
    PARTITION BY        -- (2.a)
      within_m30
    ORDER BY            -- (2.b)
      area DESC
  ) AS m30_ranking
FROM
  m30_cte
QUALIFY
  m30_ranking <= 3
ORDER BY
  m30_ranking DESC
```

name	area	within_m30	m30_ranking
Chamartín	9175481.878	true	1
Fuencarral - El Pardo	237838370.389	false	1
Villa de Vallecas	51467238.4116	false	2
Arganzuela	6462175.6297	true	2
Moncloa - Aravaca	46531136.1514	false	3
Retiro	5466211.16436	true	3

Using window functions

1. A window function **returns a value using several rows without aggregating**, but using a window instead.
2. A **window** usually consists of:
 - a **partition** criteria,
 - An **ordering** criteria,
 - A window **boundary** definition
3. To filter using window functions, use the **QUALIFY** syntax

Advanced Data Visualization Workshop: Assessing the damages of La Palma Volcano with CARTO Workflows + Builder

[Access full guide:](#)

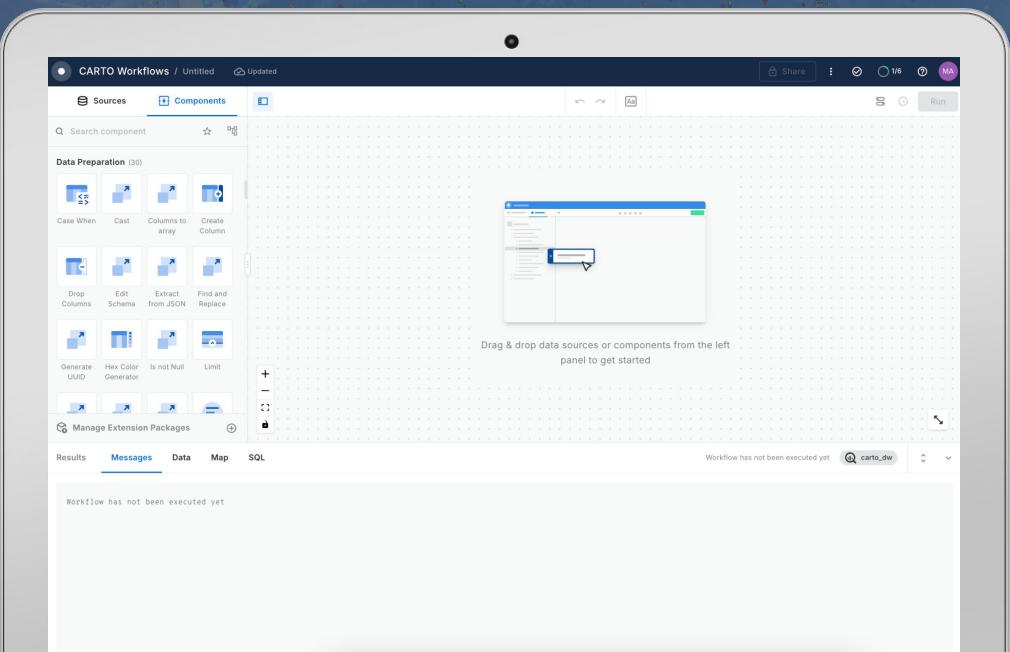
La Palma datasets

- **Lava flow:** it includes the lava flow from the Volcano eruption in La Palma, Spain as measured by the Copernicus satellite on 10/04/2021
- **Buildings:** it contains the buildings in La Palma as obtained from the Spanish cadaster website
- **Spatial Features:** provides derived variables across a wide range of themes including demographics, points of interest, and climatology data



CARTO — Unlock the power of spatial analysis

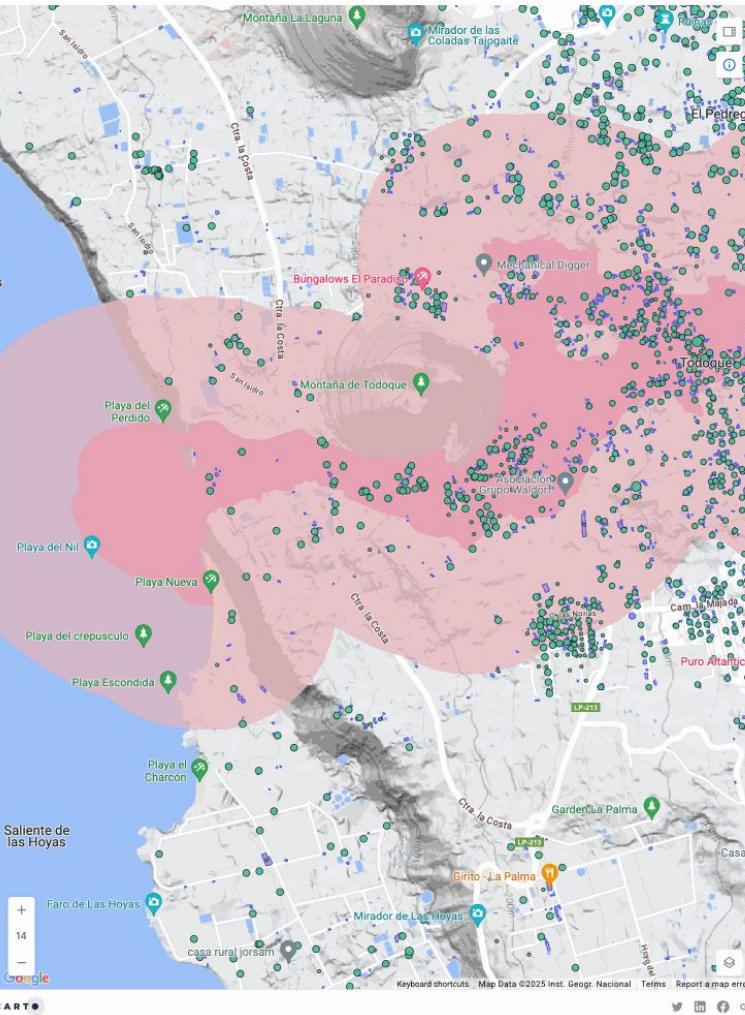
Open Lab Time



Modify & improve our La Palma Volcano map...

- Change the styling: try new palettes, change the colouring of the buildings...
- Include different widgets to extract more insights
- Show other information:
 - Visualize socio-demographic, socio-economic, and cadaster dataset aggregated at cadastral lot:

```
@ CARTODW > carto-demo-data >
demo_tables >
lapalma_sociodemo_parcels
```
 - Add a 500m-buffer around the lava flow data to show the nearby affected regions



La Palma Volcano Eruption Impact Analysis 🌋

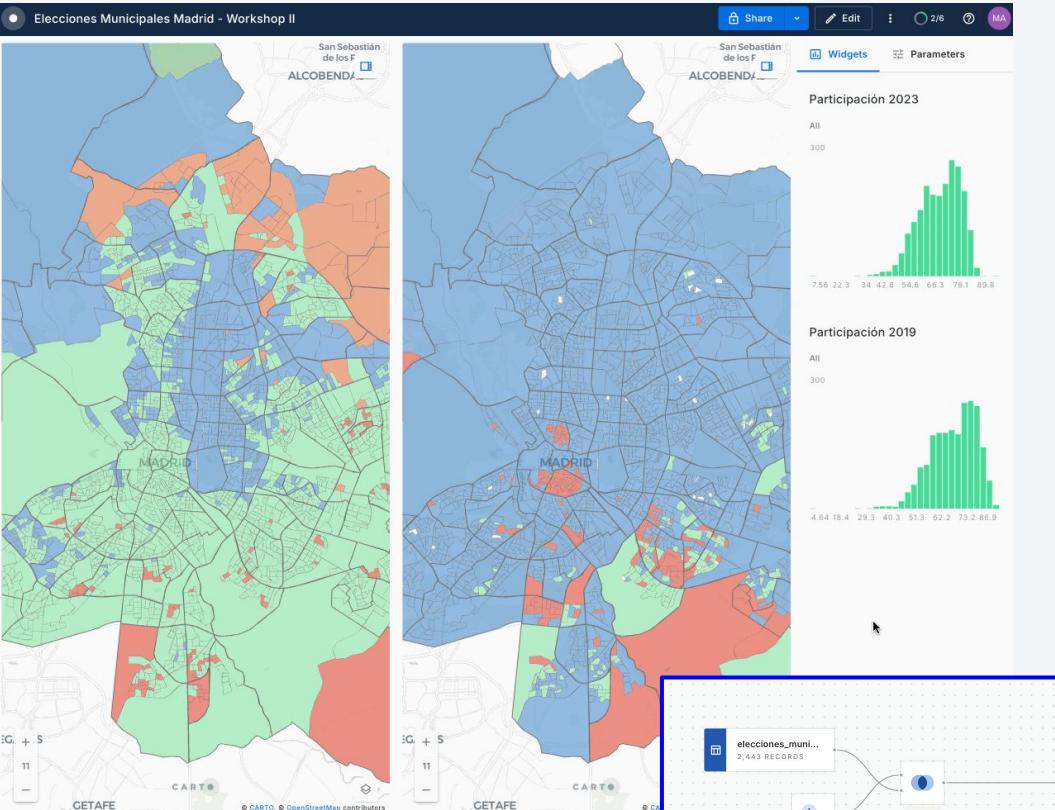


This interactive map provides an in-depth visualization of the impact caused by La Palma volcano eruption which took place in 2021. It helps understanding the extent of the eruption's effects on the local community and environment.

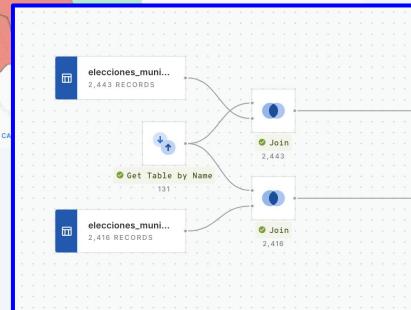
Explore the Map to Uncover:

- ⚠️ **Volcano Lava Flow Visualization:** Trace the path of the lava flow, providing a stark visualization of the affected zones.
- ● **Buffered Lava Flow Zone:** View the 500-meter buffer zone around the lava flow, marking the wider area influenced by the eruption.
- 📍 **Building and Parcel Analysis:** Investigate how buildings and sociodemographic parcels in La Palma were impacted, revealing the eruption's reach on properties and people.
- 💡 **Interactive Insights on Impact:** Engage with the lava flow areas to discover key data, such as the estimated value of affected properties, the number of properties impacted, and detailed population statistics.

Interested in Replicating This Map?
Access our tutorial in the CARTO Academy for step-by-step guidance.



```
@ CARTODW > organization data > shared_us
> elecciones_municipales_2019
> elecciones_municipales_2023
@ cartobq.docs.barrios_madrid
```



Or create a map from scratch...

Comparison of Madrid's Election Results

- Use Workflows to prepare your data and compute:
 - Right vs left-leg total votes
 - % of votes per party
 - % of participation
 - Vote variation from 2019 to 2023 per party
 - Census sections that changed winners
 - ...
- Then load it to Builder and create a visualization!

Questions?

Lucía García-Duarte
Diego Vicente
Miguel Álvarez

lgarciaduarte@carto.com
dvicente@carto.com
malvarez@carto.com

CARTO



Hudson River