

Matt Forrest · Aug 8, 2019

Stop Using Zip Codes for Geospatial Analysis

SPATIAL ANALYTICS 4 mins read Share

This post may describe functionality for an old version of CARTO. Find out about the latest and cloud-native version here.

The last time you used your zip code, you were most likely entering your address into a website to make a purchase, finding a store near your home or office, or filling out some other online form. You likely found the answer you were looking for and didn't stop to think further about that five-digit code you'd just typed out.

However, lots of companies, marketers and data analysts spend hours looking at zip codes. They are deciding how to use data tied to those zip codes to understand trends, run their businesses, and find new ways to reach you - all using that same five-digit code.

Even though there are different place associations that probably mean more to you as an individual, such as a neighborhood, street, or the block you live on, the zip code is, in many organizations, the geographic unit of choice. It is used to make major decisions for marketing, opening or closing

stores, providing services, and making decisions that can have a massive financial impact.

The problem is that zip codes are not a good representation of real human behavior, and when used in data analysis, often mask real, underlying insights, and may ultimately lead to bad outcomes. To understand why this is, we first need to understand a little more about the zip code itself.

Looking to learn more about Spatial Data Science? Check out our range of Spatial Data Science events!

The Zip Code: A Brief History

The predecessor to the zip code was the postal zone which represented a post office department for a specific city. For example:

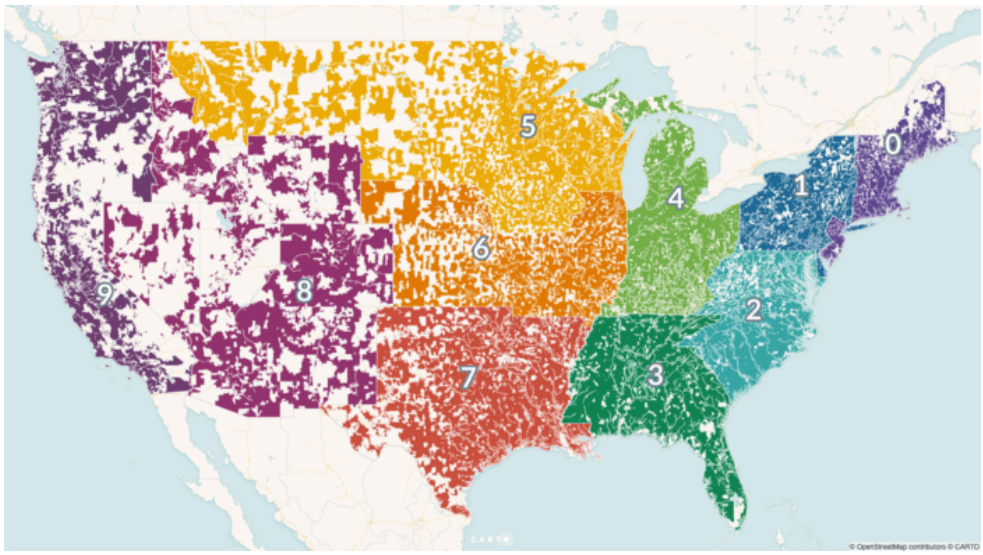
Mr. John Smith
3256 Epiphenomenal Avenue
Minneapolis 16 Minnesota

"16" represents the postal zone in a Minneapolis. But with more and more mail being sent, in 1963 the Postal Service decided to roll out the **Zone Improvement Plan** which transformed addresses to look like the following:

Mr. John Smith
3256 Epiphenomenal Avenue
Minneapolis MN 55416

The five digit code represents a part of the country (5_ _ _ _) a sectional center facility (_ 5 4 _ _) and the associate post office or delivery area (_ _ _ 1 6).

H2 Intellect - Exhibit 2041
Home Depot v. H2 Intellect - IPR2025-00480
Page 2 of 20



The first digit for every zip code for the states in the contiguous United States

By 1967 ZIP codes were made mandatory for bulk mailers and continued to be adopted by almost anyone sending mail in the US. Over time, the ZIP+4 was added to add more granularity to the zip code to denote specific locations even buildings for postal workers to deliver. The Postal Service even created a character, Mr. Zip, to promote the use of ZIP codes, who was featured on stamps, commercials and songs.



H2 Intellect - Exhibit 2041
Home Depot v. H2 Intellect - IPR2025-00480
Page 3 of 20

ZIP codes themselves do not actually represent an area, rather a collection of routes:

Despite the geographic derivation of most ZIP Codes, the codes themselves do not represent geographic regions; in general, they correspond to address groups or delivery routes. As a consequence ZIP Code "areas" can overlap be subsets of each other, or be artificial constructs with no geographic area (such as 095 for mail to the Navy which is not geographically fixed). In similar fashion, in areas without regular postal routes (rural route areas) or no mail delivery (undeveloped areas) ZIP Codes are not assigned or are based on sparse delivery routes and hence the boundary between ZIP Code areas is undefined.

The US Census provides data for ZIP Code Tabulation Areas, or geographic files:

ZIP Code Tabulation Areas (ZCTAs) are generalized areal representations of United States Postal Service (USPS) ZIP Code service areas.The USPS ZIP Codes identify the individual post office or metropolitan area delivery station associated with mailing addresses. USPS ZIP Codes are not areal features but a collection of mail delivery routes.

Here we find our first problem with ZIP Codes, that they do not represent an actual area on a map, but rather a collection of routes that help postal workers effectively deliver mail. They aren't designed to measure sociodemographic trends as a business would generally want to do. You can actually look up individual delivery routes like the one below:

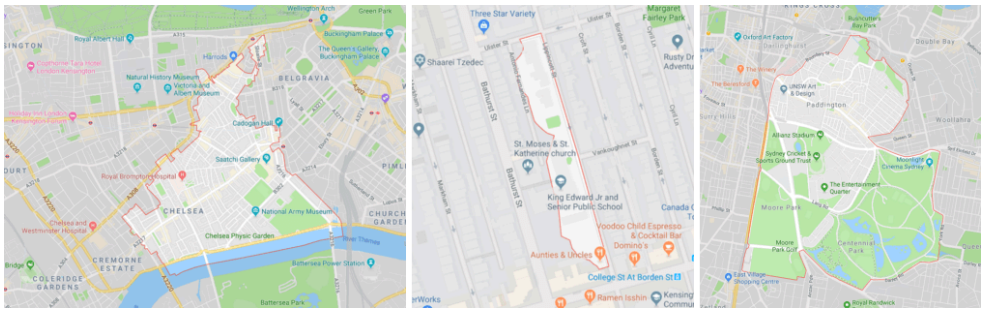


H2 Intellect - Exhibit 2041
Home Depot v. H2 Intellect - IPR2025-00480
Page 4 of 20



One zip code route in New York's East Village

We are only scratching the surface of the issue here. Similar issues exist around the world, with postal codes representing strange boundaries



Postal Codes in London, Toronto, and Sydney.

Using ZIP codes for data analysis

Fast forward to today, where many companies can easily look into their database and find a dataset with a zip_code column in it,

H2 Intellect - Exhibit 2041
Home Depot v. H2 Intellect - IPR2025-00480
Page 5 of 20

which allows them to group and aggregate data to see trends and business performance metrics. As stated earlier, the problem with ZIP Codes is that:

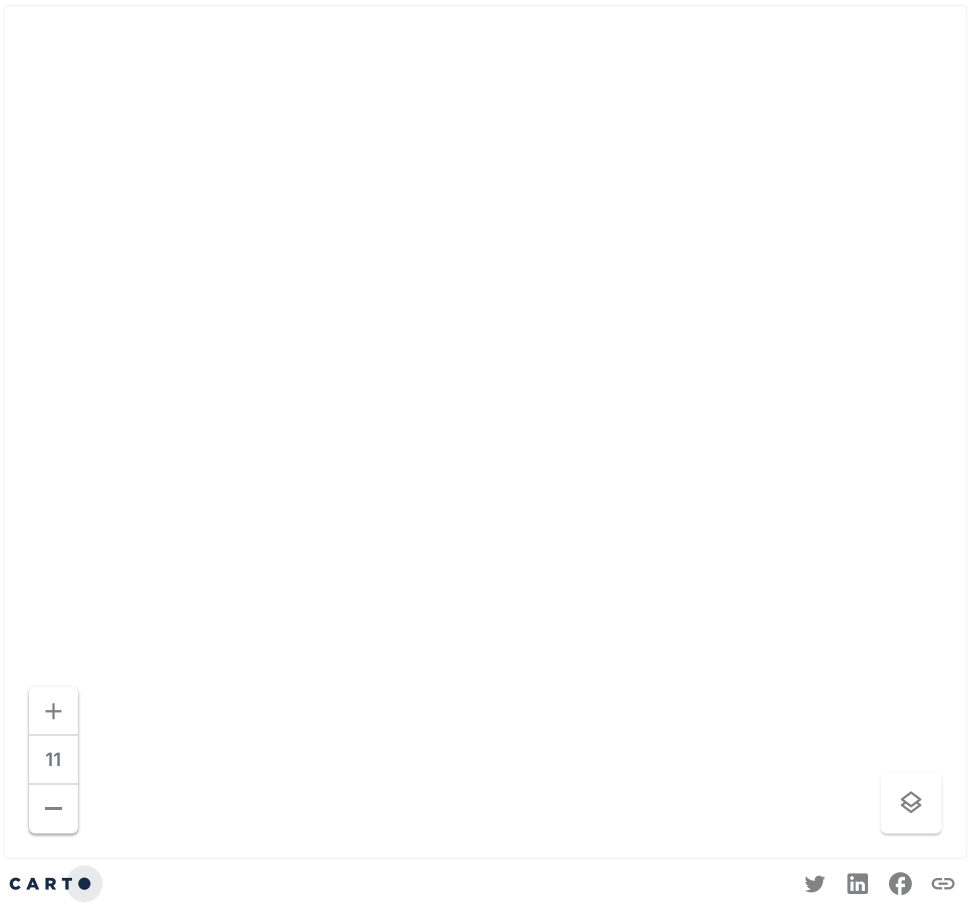
1. They don't represent real boundaries but rather routes
2. They don't represent how humans behave

The later represents two specific issues in using spatial data: spatial scale of observations and spatial scale support (you can learn more about this in this lecture from UChicago's Luc Anselin, here). The first is that humans don't behave based on administrative units such as zip codes, or even census units. Their behavior is influenced much more by their neighbors, or areas such as a neighborhoods or high activity areas (such as central business districts). The second is that spatial data is provided at multiple scales, and many times those boundaries are overlapping or nested within another boundary.

Let's look at an example of this in one specific area in Dallas.



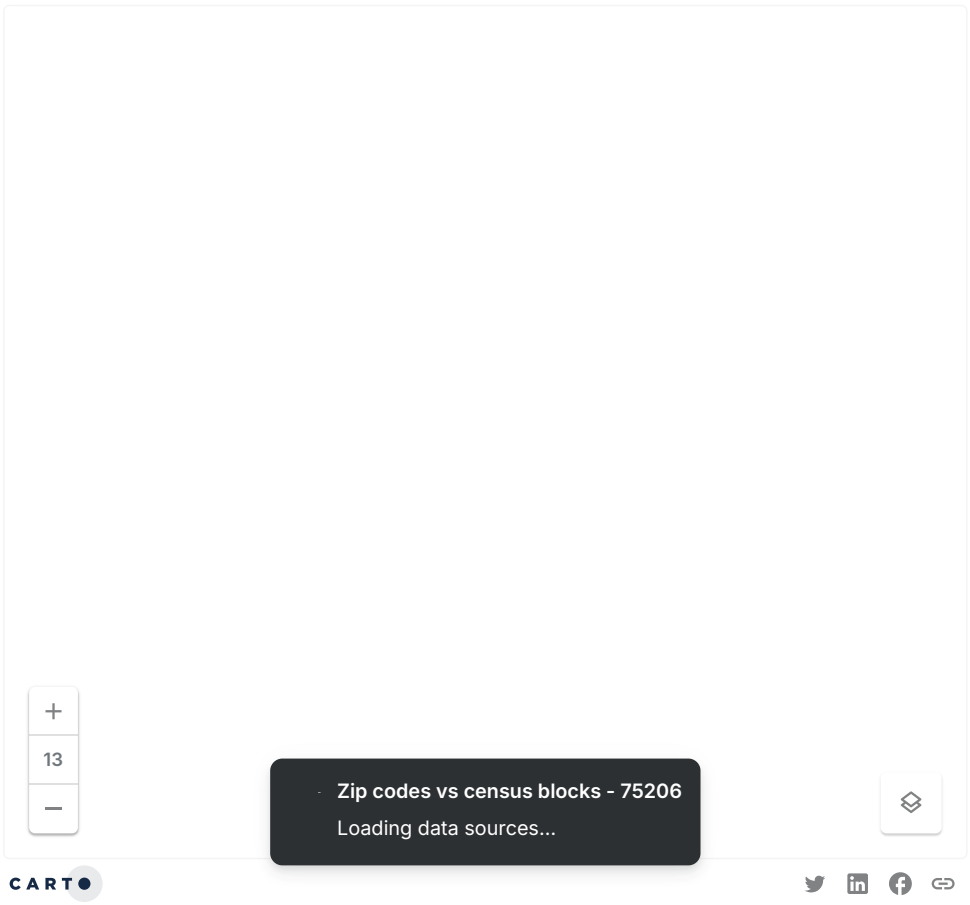
H2 Intellect - Exhibit 2041
Home Depot v. H2 Intellect - IPR2025-00480
Page 6 of 20



In this map, we can see large white boundaries, which represent ZIP code boundaries, and below them are boundaries for US Census Block Groups. The darker green represents higher income, as provided by the US Census. You can access both of these datasets - with associated sociodemographic data - via our Spatial Data Catalog.

When we look at one specific ZIP Code we can see that income data in even more detail:

H2 Intellect - Exhibit 2041
Home Depot v. H2 Intellect - IPR2025-00480
Page 7 of 20



What we can see is that 12 month median household income in this single zip code (75206) ranges from $9,700 to $227,000 when we look at block groups that completely or partially fall within this single ZIP Code, which the Census lists as having a median household income of $63,392.

Median Income is one way to evaluate the range of values within a ZIP Code (keep in mind these are ZCTA boundaries) but we can likely see variance like this in population, employment, and other relevant metrics for data analysis.

Sticking with median household income, we decided to expand this analysis to the entire United States, to see which areas are the least and most in-equal when you look at ZIP Codes and the Census Block Groups that intersect with the ZCTA Boundaries.

H2 Intellect - Exhibit 2041
Home Depot v. H2 Intellect - IPR2025-00480
Page 8 of 20

The most unequal zip code is 33139 in Miami Beach FL

1. 33139: Miami FL ($241,344 Difference)
2. 44120: Cleveland OH ($237,501 Difference)
3. 10013: New York NY ($233,559 Difference)
4. 10023: New York NY ($233,157 Difference)
5. 11201: Brooklyn NY ($233,031 Difference)
6. 10601: White Plains NY ($232,813 Difference)
7. 33141: Miami FL ($232,633 Difference)
8. 92648: Huntington Beach CA ($231,290 Difference)
9. 98105: Seattle WA ($230,906 Difference)
0. 33143: Miami FL ($230,626 Difference)

The most similar* zip code is in Chesapeake WV

1. 25315: Chesapeake WV ($2 Difference)
2. 79357: Cone TX ($4 Difference)
3. 65052: Linn Creek MO ($9 Difference)
4. 73093: Washington OK ($12 Difference)
5. 68370: Hebron NE ($13 Difference)
6. 19541: Mohrsville PA ($15 Difference)
7. 05340: Bondville VT ($18 Difference)
8. 12958: Mooers NY ($26 Difference)
9. 19941: Ellendale DE ($37 Difference)
0. 54896: Loretta WI ($38 Difference)

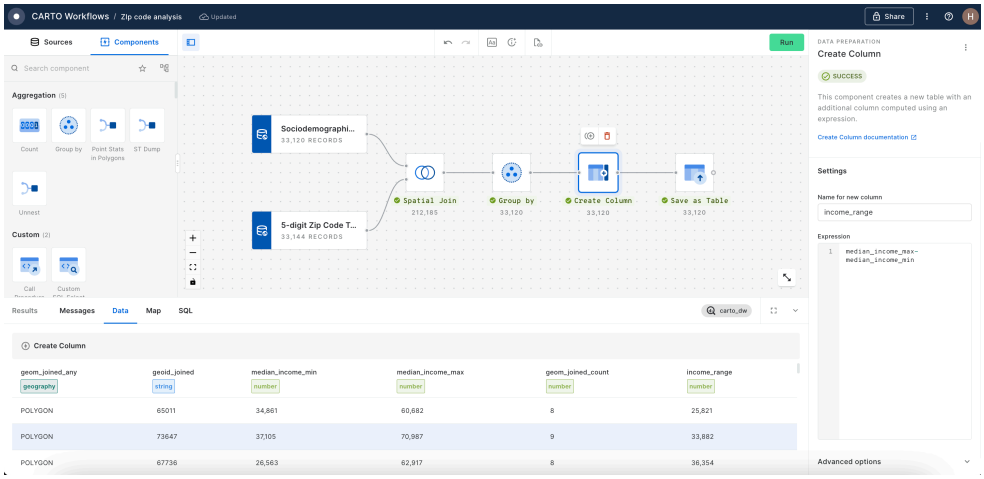
** similar where the difference is greater than 0*



H2 Intellect - Exhibit 2041
Home Depot v. H2 Intellect - IPR2025-00480
Page 9 of 20

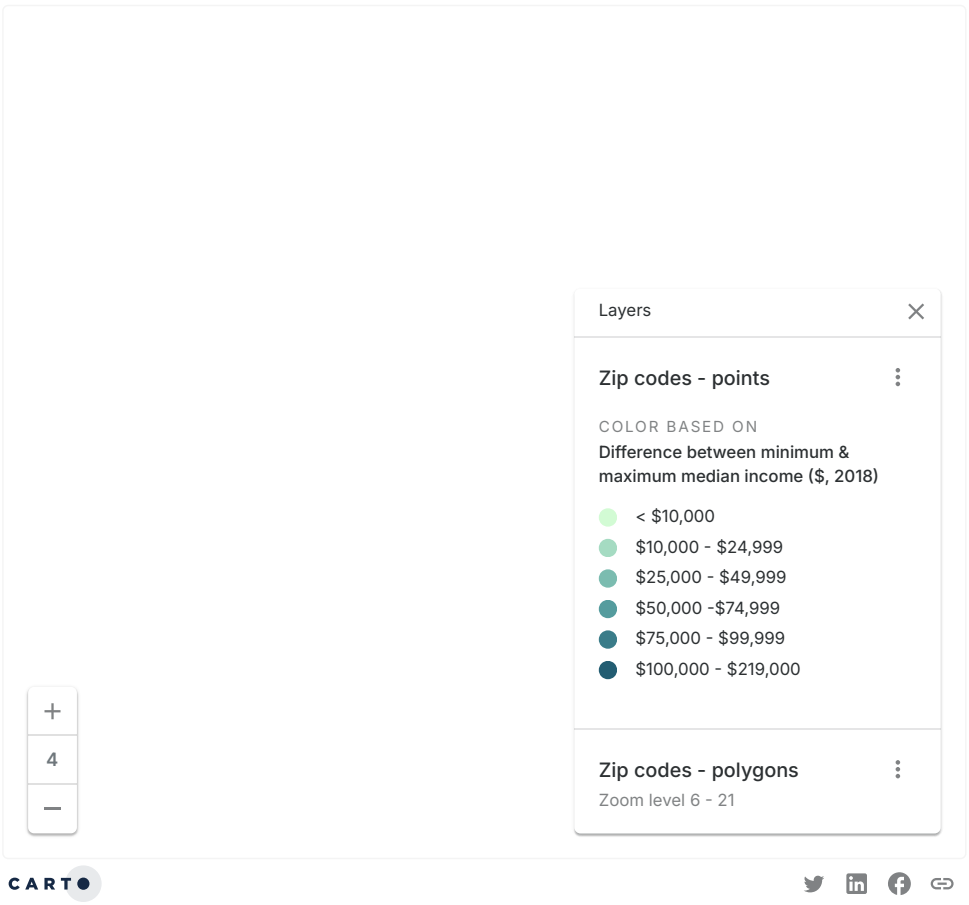
We can quickly use CARTO Workflows to perform this analysis for all of the United States without code. If you'd like to have a go yourself, you can sign up for a free 14-day CARTO trial here.

In this workflow (see below), we used a Spatial Join to establish which census blocks intersect each Zip Code. Next, the Group By component was used to calculate the minimum and maximum income for each zip code, after which the Create Column component was used to calculate the difference between the two.



After creating this as a new table we can see that the majority of the most unequal zip codes tend to be in cities or larger metro areas and more equal zip codes tend to be in rural areas around the country.

H2 Intellect - Exhibit 2041
Home Depot v. H2 Intellect - IPR2025-00480
Page 10 of 20



Check out the full map.

Sign me up for a FREE account

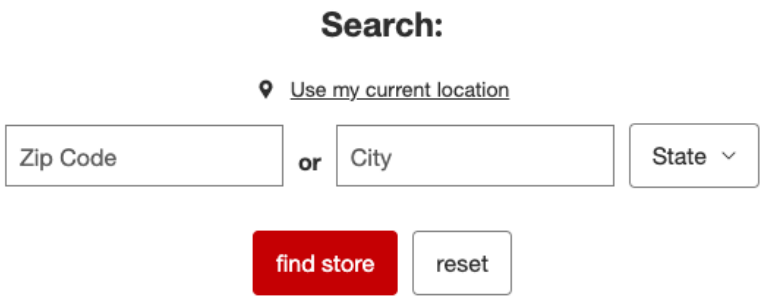
Learn more about how organisations are changing the way they perform spatial analysis with our FREE ebook The State of Spatial Data Science 2024!



H2 Intellect - Exhibit 2041
Home Depot v. H2 Intellect - IPR2025-00480
Page 11 of 20

So why do we use ZIP Codes?

In practice, it is easy to use zip codes as a geospatial unit. As we stated earlier, almost any e-commerce or delivery service, or app that uses location or needs to locate their users in any way will collect a zip code. Additionally everyone is familiar with zip codes as they are part of any address in most countries.



'Find A Store' on Target.com

Terms like Census Block Group or Tract are less familiar to those who don't work with geospatial data on a regular basis and they can be more difficult to find and harder to work with especially if you aren't familiar with terms like Shapefile, FTP, and ETL. Even then you have to go through the US Census FTP website,, download geographic files state by state and join those files to census measures.

Finally, most people know without looking at a map that zip codes represent a smaller area than a city, but larger than say a neighborhood. Conceptually, they feel small enough to get a very focused view of the world, and big enough to capture enough of a sample size of trends.

H2 Intellect - Exhibit 2041
Home Depot v. H2 Intellect - IPR2025-00480
Page 12 of 20

The short answer is that zip codes are easy to find, familiar, and provide a granular enough view of the world (or so we thought).

With that said there are real world problems that arise from using zip codes in geospatial analysis. One example is in real estate where, in many cities or areas, there are homes listed in a zip code even though as we know those boundaries are arbitrary. This article from the Harvard Business Review also describes a similar phenomenon with Airbnb listings and rising rent prices.

In simple terms we argue that if a zip code is "touristy," meaning it has a lot of restaurants and bars, and if awareness of Airbnb increases which we measure using the Google search index for the keyword "Airbnb," then any jump in Airbnb supply in that zip code is likely driven by an increase in demand for short-term rentals through Airbnb rather than local economic conditions.

More importantly, using zip codes for analysis can mask serious conditions that are taking place at a different spatial scale. The Flint water crisis was one of these cases. This article by Richard Casey Sadler, an Assistant Professor at Michigan State University describes the problem in great detail and raises similar points about zip codes (the full article is well worth reading):

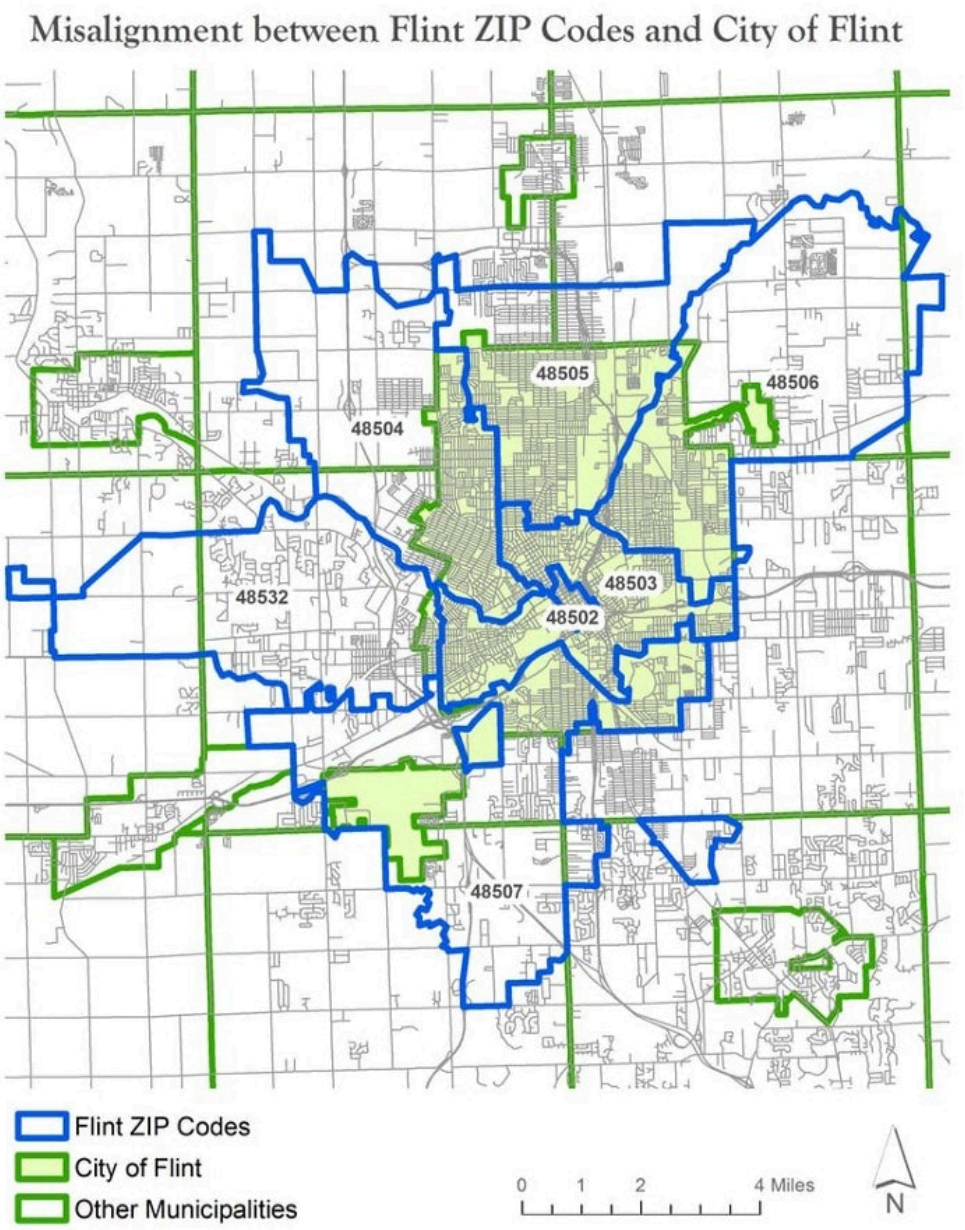
Dr. Tony Grubesic, an Arizona State University professor has called them "one of the quirkier 'geographies' in the world." Dr. Nancy Krieger, a Harvard University professor and colleagues have called out their unacceptability for small-area analyses.

Ultimately the state used zip codes to analyze the blood lead statistics in aggregate, which effectively masked the actual issue. This is because:

In Flint's case, the state's error was introduced because "Flint ZIP codes" do not align well with the city of Flint or its water system.

H2 Intellect - Exhibit 2041
Home Depot v. H2 Intellect - IPR2025-00480
Page 13 of 20

The city and water system are almost 100 percent coterminous – that is they share the same boundaries... In total, Flint ZIP codes used in the state's analysis blanket parts of eight different municipalities (seven townships and one city) surrounding Flint.



Map by Richard Casey Sadler

H2 Intellect - Exhibit 2041
Home Depot v. H2 Intellect - IPR2025-00480
Page 14 of 20

In the case of Flint simply looking at a different spatial scale or analysis may have shown the problem more clearly. For fields like public health and critical services understanding and using appropriate spatial scale is critically important.

What else can we do?

So if you are compelled to do away with zip code analysis, the good news is that there are several different options available.

Use Addresses

The first and foremost recommendation would be to use real world addresses. When you know an address string, you can use a geocoder - or the service that Google or other map search engines use to take an address and transform it into latitude and longitude. Most every major service offers an API and wrappers for different languages to do this. Keep in mind there are generally some best practices for cleaning your data and you will need valid address strings to do so. Other tools like Libpostal can help you normalize and parse your address strings.

Use Census Units

You can also use Census units such as a Census Block Group or Tract. As I mentioned it has not always been easy to find and collect this data at scale but there are many new tools that are becoming available to use. CenPy is a Python library that allows you to connect and find Census data (good tutorial here) where you can find measures from the decennial census or American Community Survey. CARTO also provides Census and ACS through the Data Observatory which was used in the notebook for the full US analysis.

H2 Intellect - Exhibit 2041
Home Depot v. H2 Intellect - IPR2025-00480
Page 15 of 20

You can also find out Census geometry IDs for a specific address location using the US Census Geocoder. You can pass in a variety of parameters via the API or use it in Python with the censusgeocode package.

As your gather address data, you can easily add in a Census Tract or Block Group ID to that entry and use that rather than the zip code field in your data. This will allow you to do the same aggregation you were doing except at a more appropriate geographic scale.

Use your own Spatial Index

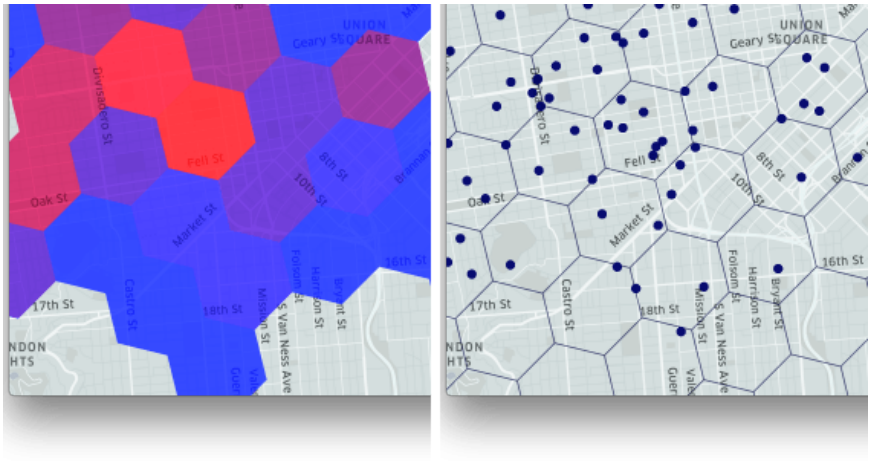
Finally - and arguably the best option - is to use a Spatial Index.

Spatial Indexes like H3 and quadkey are global, multi-resolution grid systems which are geolocated by a short reference ID, rather than a complex geometry. This makes them super lightweight to store and lightning-fast to analyze. The grid cells have the same shape (although their size varies by latitude), meaning data is evenly geographically distributed - avoiding boundary and visual bias which is often caused by irregular grids like zip code; read more here.

A key benefit here is the anonymization of sensitive data. Given that address data can be sensitive you can create a data pipeline that simply reads incoming addresses geocodes them, assigns a spatial index, then passes that indexed data into a separate table, then you can store or delete the address data as needed.



H2 Intellect - Exhibit 2041
Home Depot v. H2 Intellect - IPR2025-00480
Page 16 of 20



Want to learn more about Spatial Indexes? Check out our ebook: Spatial Indexes 101!

Working with spatial data can be difficult but the availability of data and tools has made it more accessible. By using the correct spatial scale and discarding analysis at the zip code level you can improve the quality of your insights and create more meaningful outcomes and analysis.

Want to see how it might work with your data?

Hear from our experts!

Request a demo

Don't forget to share this post on Twitter, Facebook and Linkedin!

H2 Intellect - Exhibit 2041
Home Depot v. H2 Intellect - IPR2025-00480
Page 17 of 20

About the author

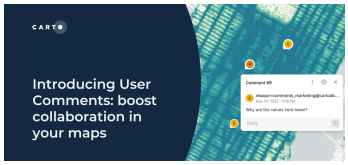


Matt Forrest is the Field CTO at CARTO and has been working with geospatial technology and data for the past 10+ years. He is passionate about all things geospatial, with a mission to enable others to use modern GIS and spatial SQL to help improve their work and advance their careers.

More Posts from Matt Forrest

Related Posts







Helen McKenzie and Ernesto Martínez Becerra • May 20, 2025

Live maps with automated Workflows integrations are here

Create live, always-updated maps with CARTO by integrating Workflows and Builder. Automate geospatial

Ana Manzanares and Helen McKenzie • May 13, 2025

Introducing User Comments: boost collaboration in your maps

Boost team collaboration with User Comments in CARTO Builder—add, track, and resolve feedback

Javier de la Torre and Helen McKenzie • Mar 13, 2025

Seamlessly integrate Google Earth Engine with CARTO Workflows

Seamlessly integrate Google Earth Engine with CARTO Workflows to analyze spatial data at scale - no coding

H2 Intellect - Exhibit 2041
Home Depot v. H2 Intellect - IPR2025-00480
Page 18 of 20

analysis to visualization in one step.

SPATIAL ANALYTICS

directly on your enterprise maps.

SPATIAL ANALYTICS

required. Unlock powerful geospatial insights today!

SPATIAL ANALYTICS

Ready to optimize your territories with Location Intelligence?

Request live demo

Pricing

Events

Webinars

Partners

Unlock the power of spatial analysis.

© CARTO 2025

Terms Privacy Notice

H2 Intellect - Exhibit 2041
Home Depot v. H2 Intellect - IPR2025-00480
Page 20 of 20