CARTO

2019

Spatial Data Science Methods for Improving Models

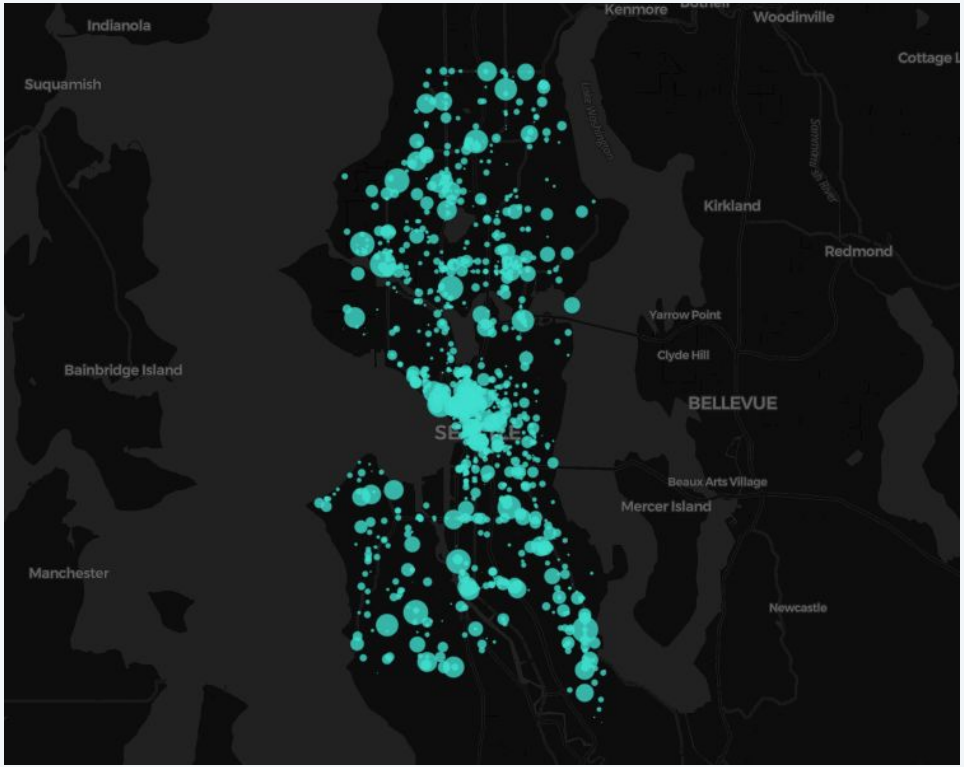
Andy Eschbacher
Data Scientist
@MrEPhysics


Overview of
Spatial Data

Points on a map

The most common way we see spatial data:

`Lat/Lng/Attributes`



Map by Mamata Akella

Zip Codes

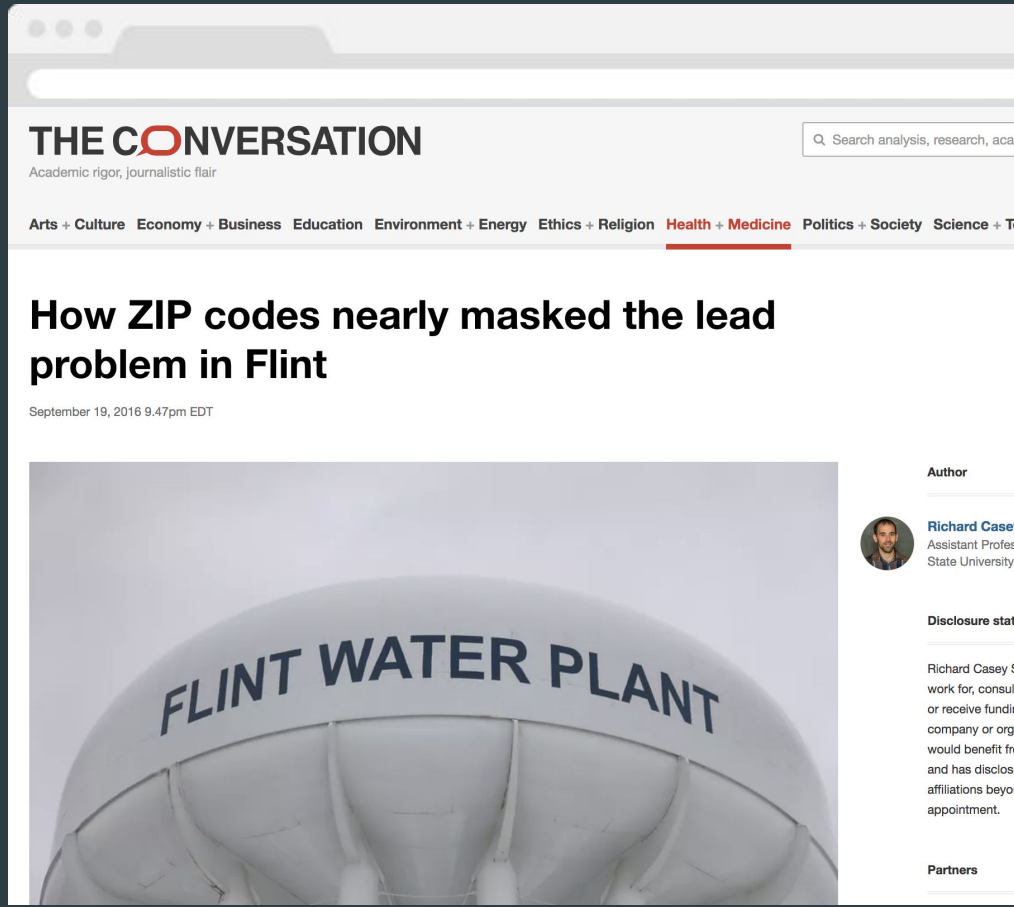
Some tips:

Try not to use zip codes.

Zip codes are not polygons. They are more akin to postal routes (lines)

Zip code != ZCTA from census

Zip codes change frequently



Real world problems

Check your unit of analysis. Zip codes don't obey city boundaries. Or really any boundaries at all.

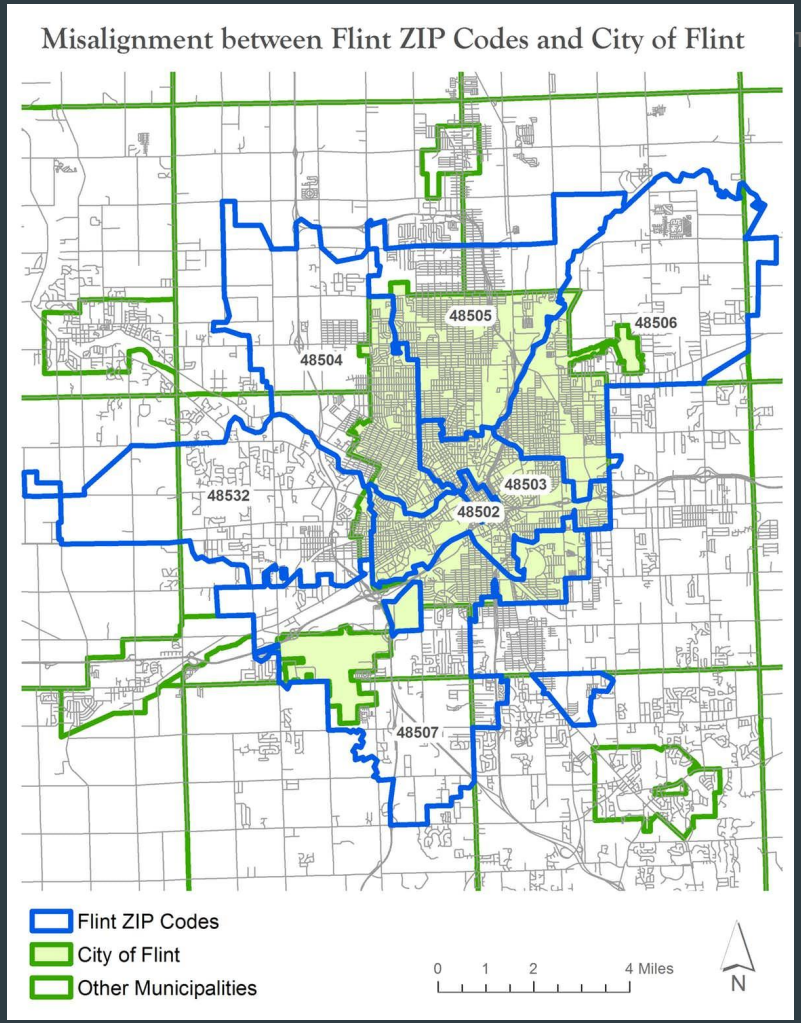
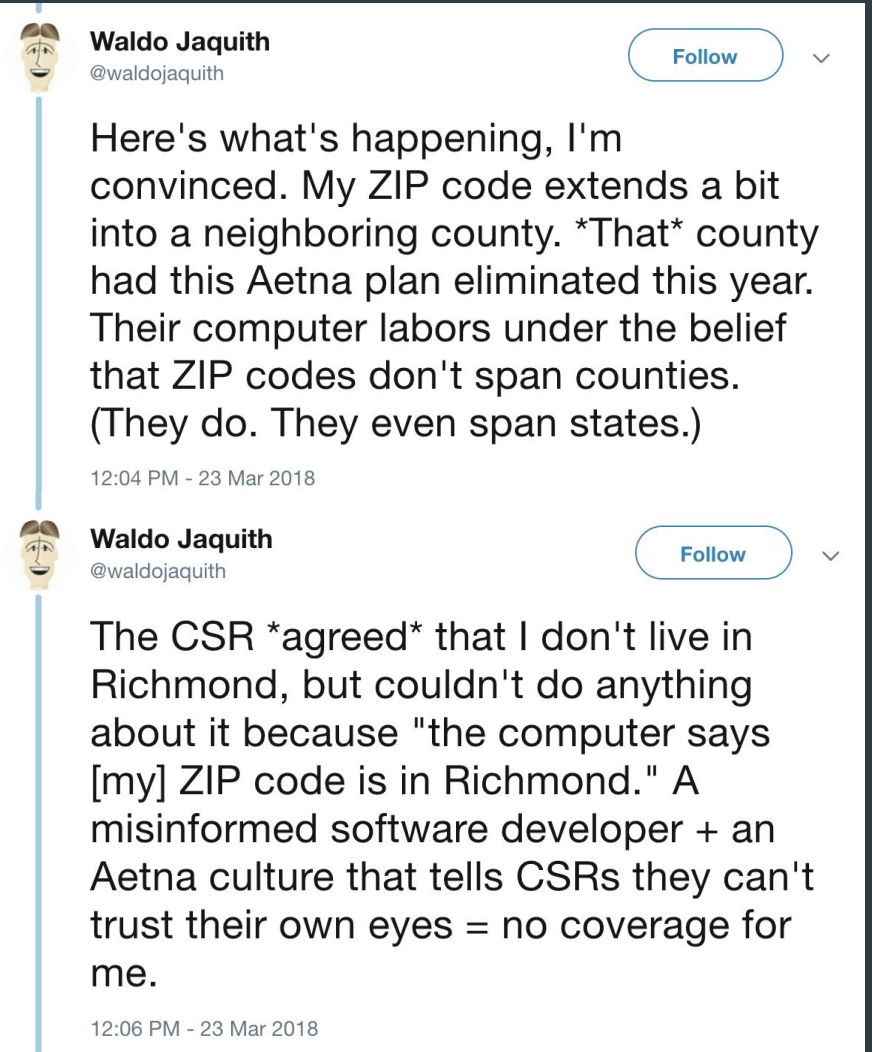


Figure from Grubesic et al

Real world problems

Another Tip:

Don't use zip codes when you are an insurance company that has plans that span a different type of geography (counties).



Waldo Jaquith
@waldojaquith
Follow

Here's what's happening, I'm convinced. My ZIP code extends a bit into a neighboring county. *That* county had this Aetna plan eliminated this year. Their computer labors under the belief that ZIP codes don't span counties. (They do. They even span states.)

12:04 PM - 23 Mar 2018

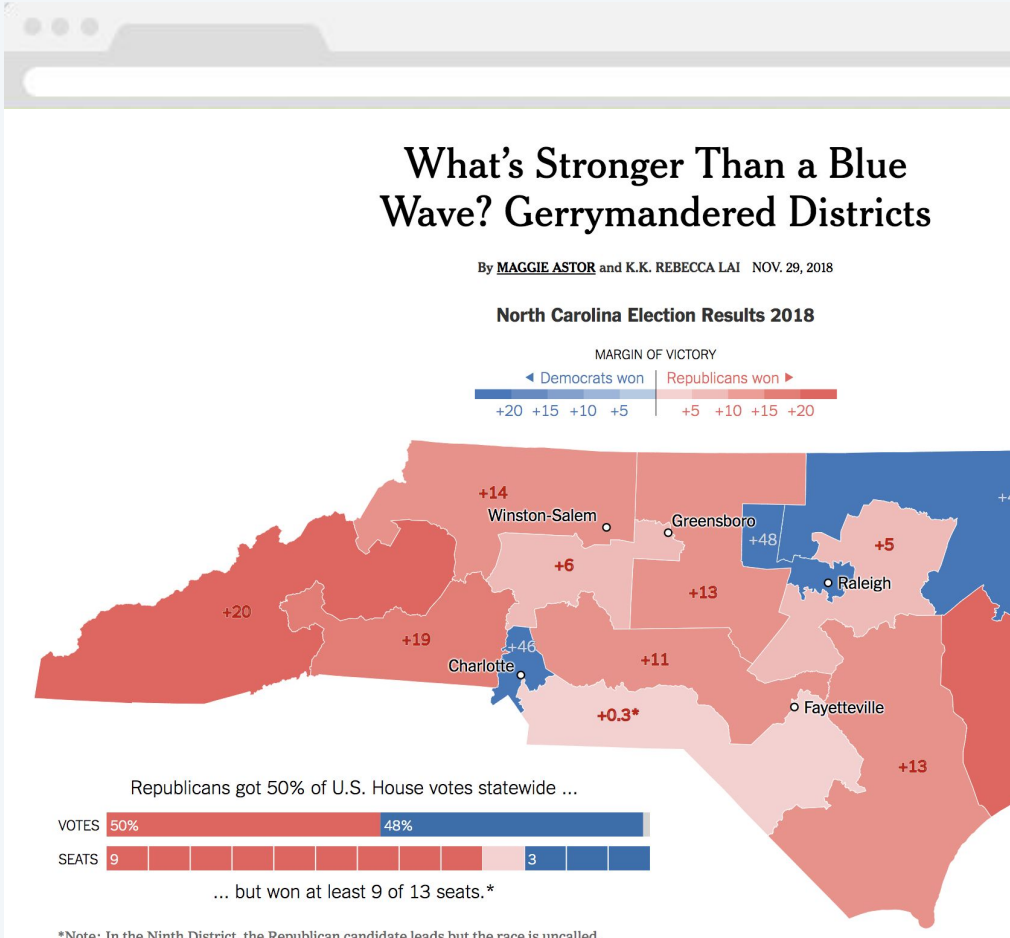
Waldo Jaquith
@waldojaquith
Follow

The CSR *agreed* that I don't live in Richmond, but couldn't do anything about it because "the computer says [my] ZIP code is in Richmond." A misinformed software developer + an Aetna culture that tells CSRs they can't trust their own eyes = no coverage for me.

12:06 PM - 23 Mar 2018

Boundaries are manipulated

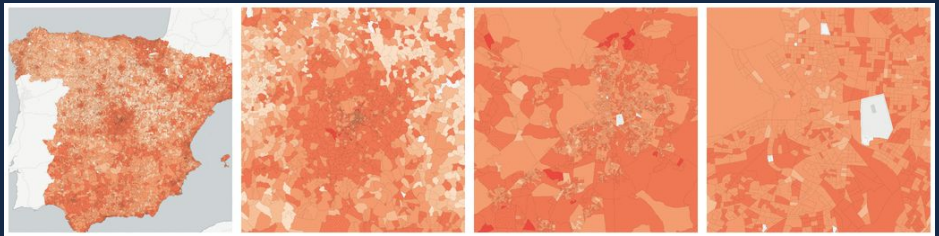
Gerrymandering shapes polygons to favor one group over another

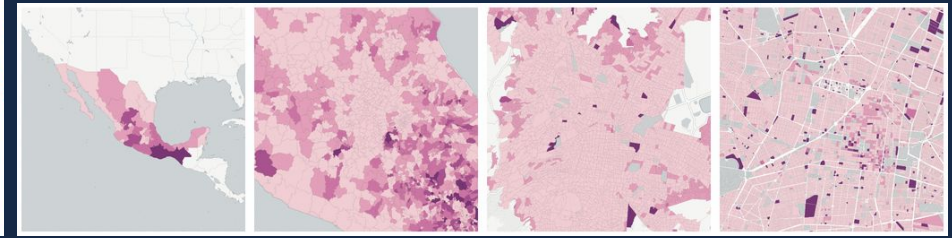


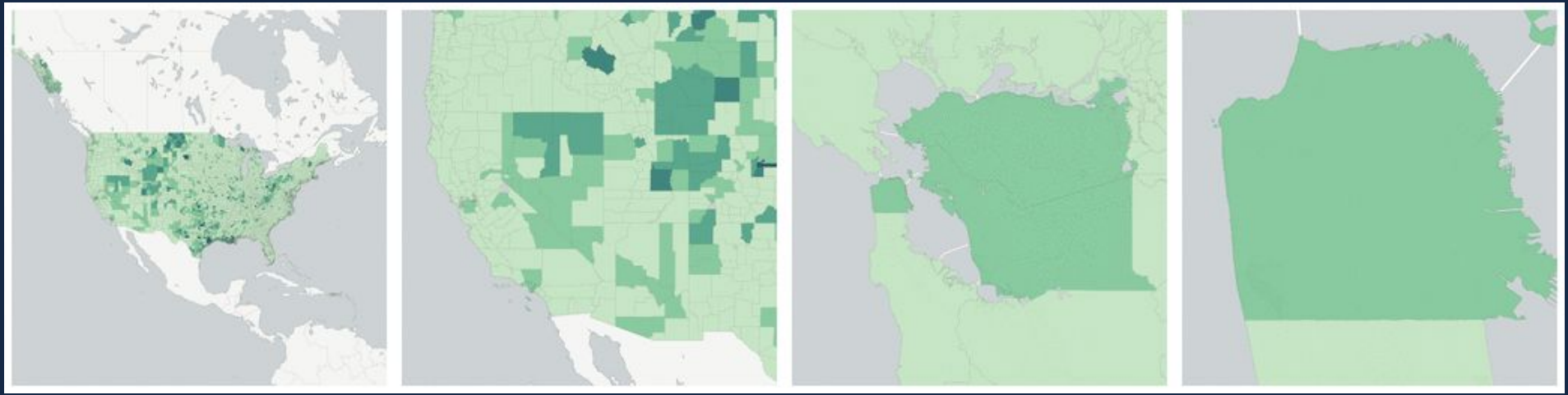
New York Times, Nov 2018

Census

Freely accessible demographic data and more at multiple geographical scales for many countries across the world





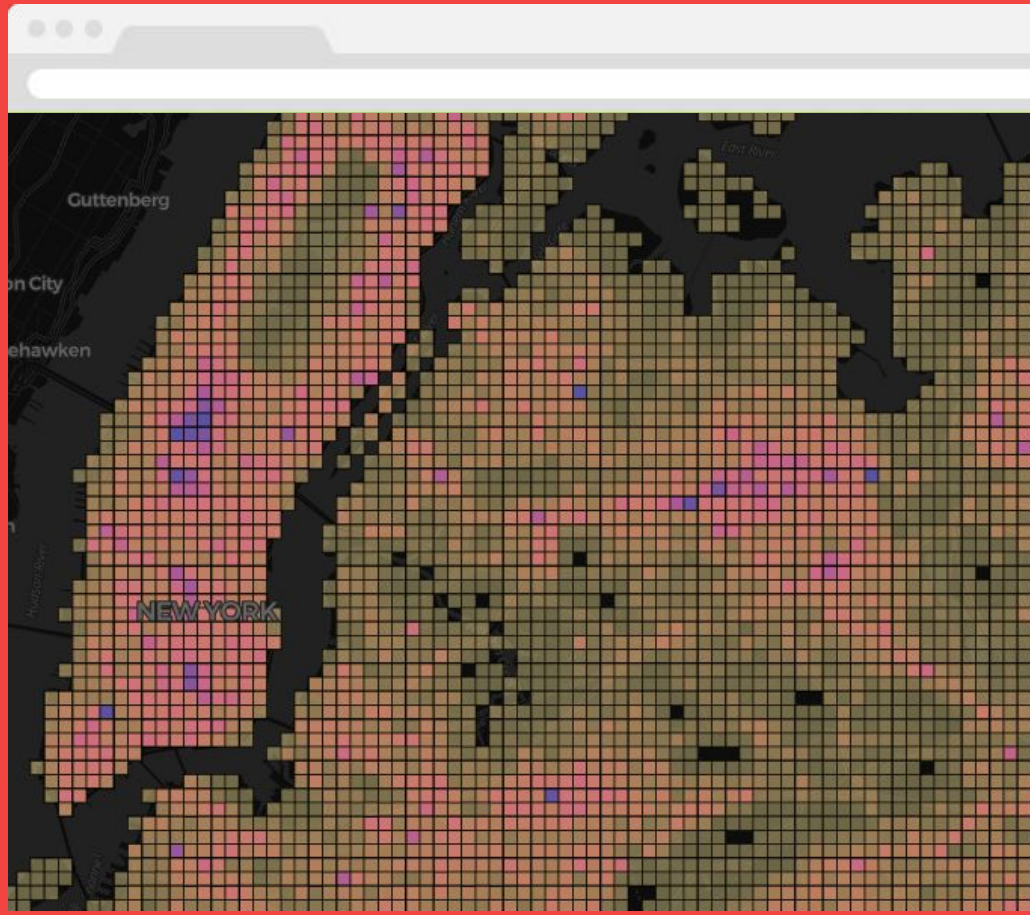


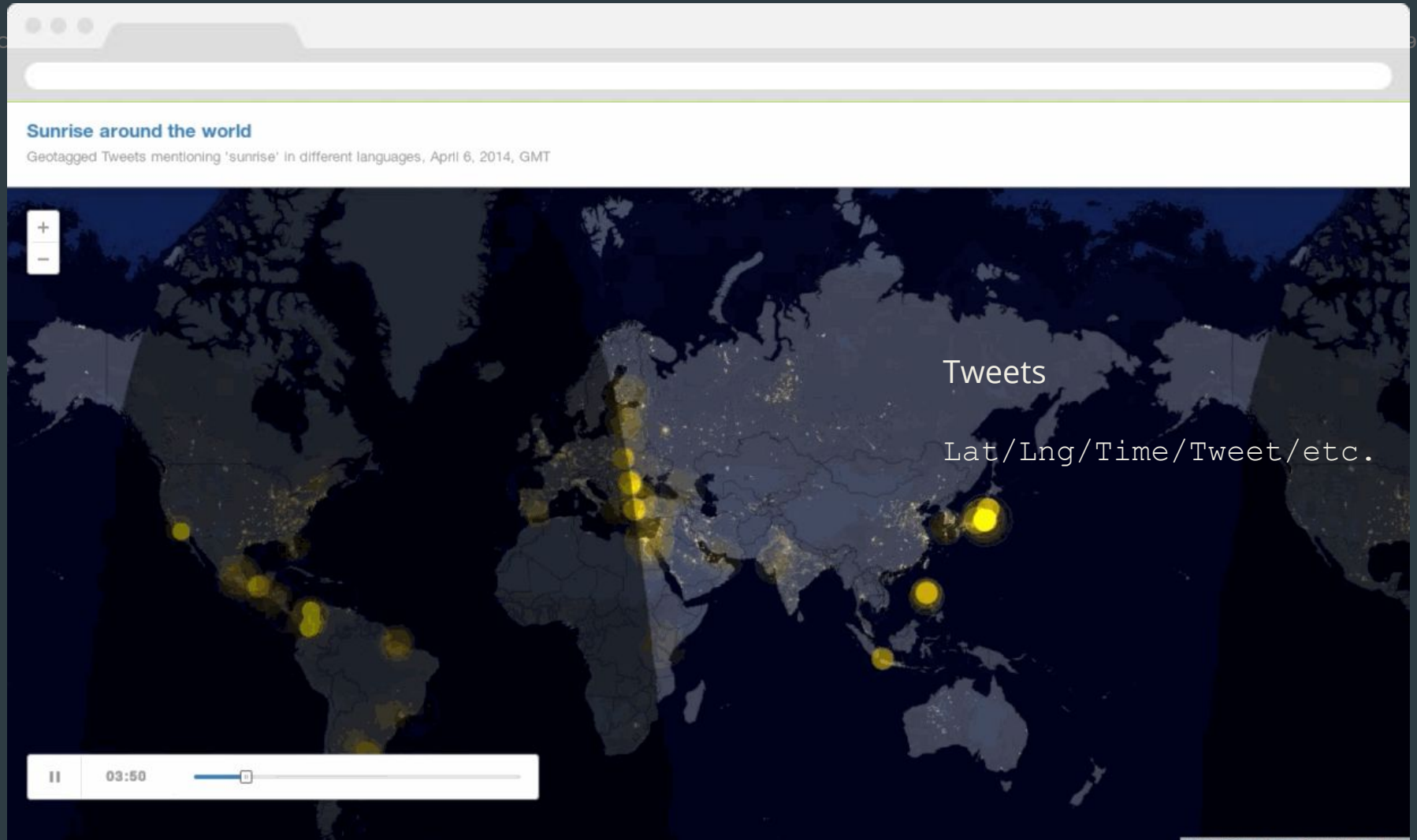
LODES

Origin-Destination Data

Modern Spatial Data Sources

Newer data sources have come about because of changes in technology




Sunrise around the world
Geotagged Tweets mentioning 'sunrise' in different languages, April 6, 2014, GMT
+
−
Tweets
Lat/Lng/Time/Tweet/etc.
03:50

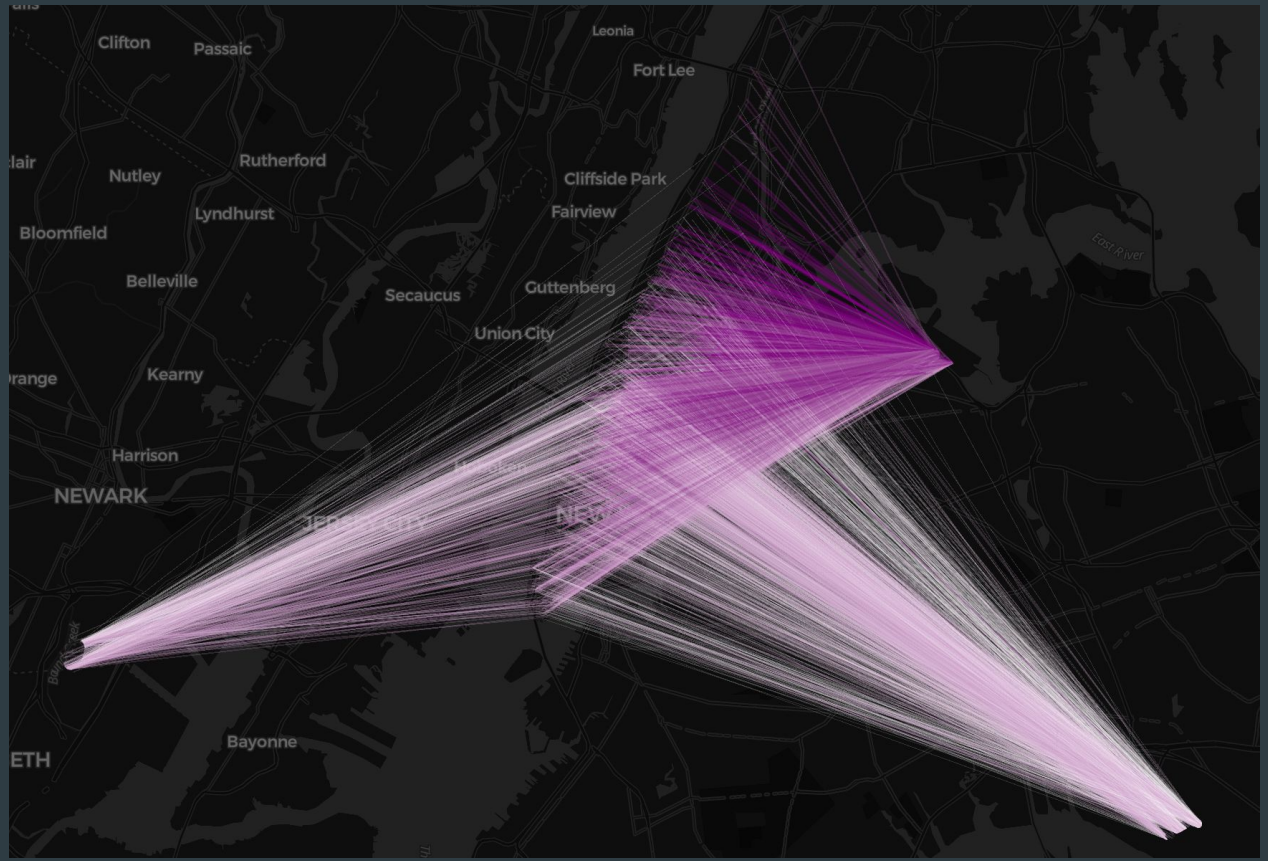
GPS Data

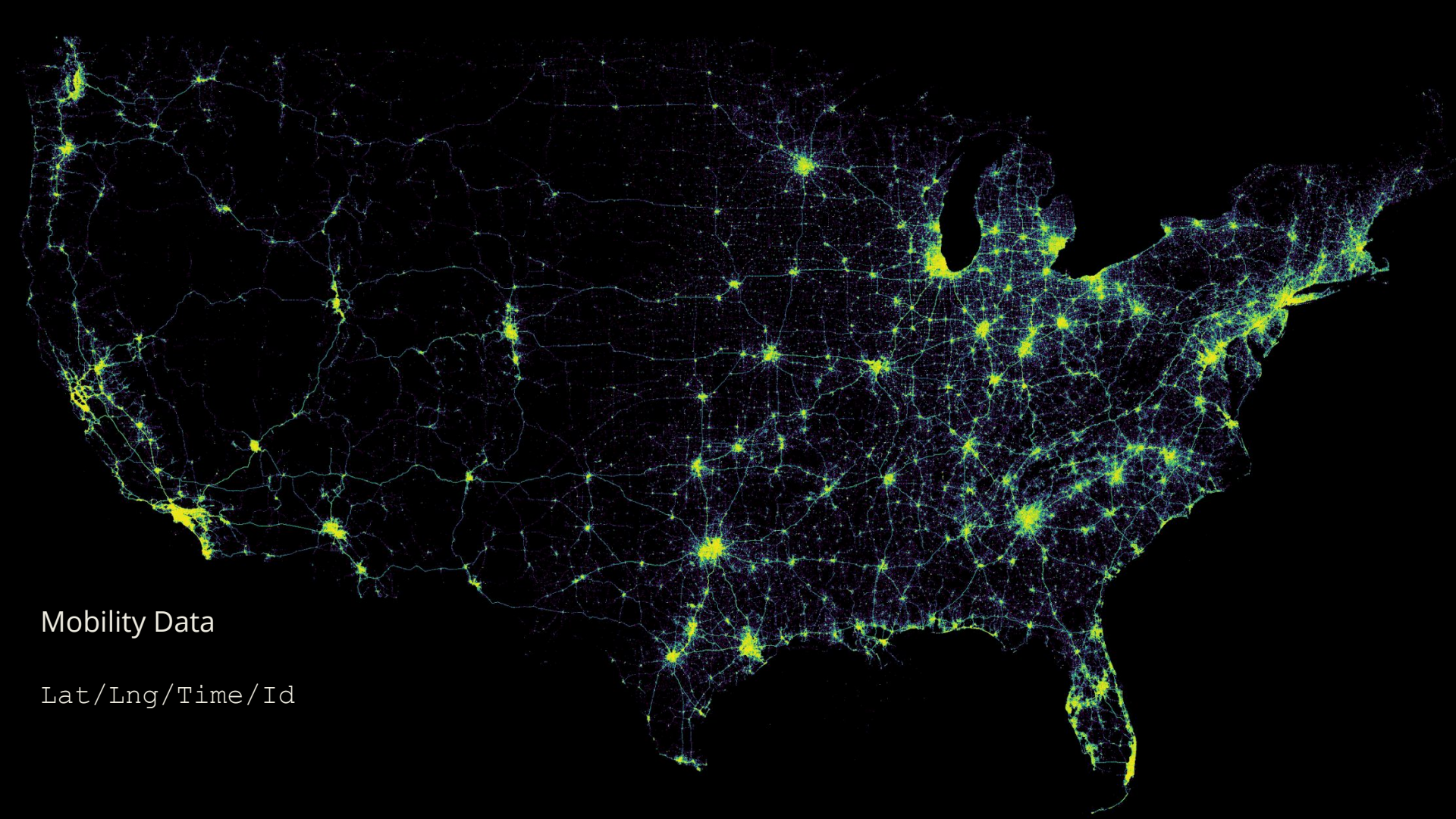
Spencer the Cat

`Lat/Lng/Time`

Open Taxi Data

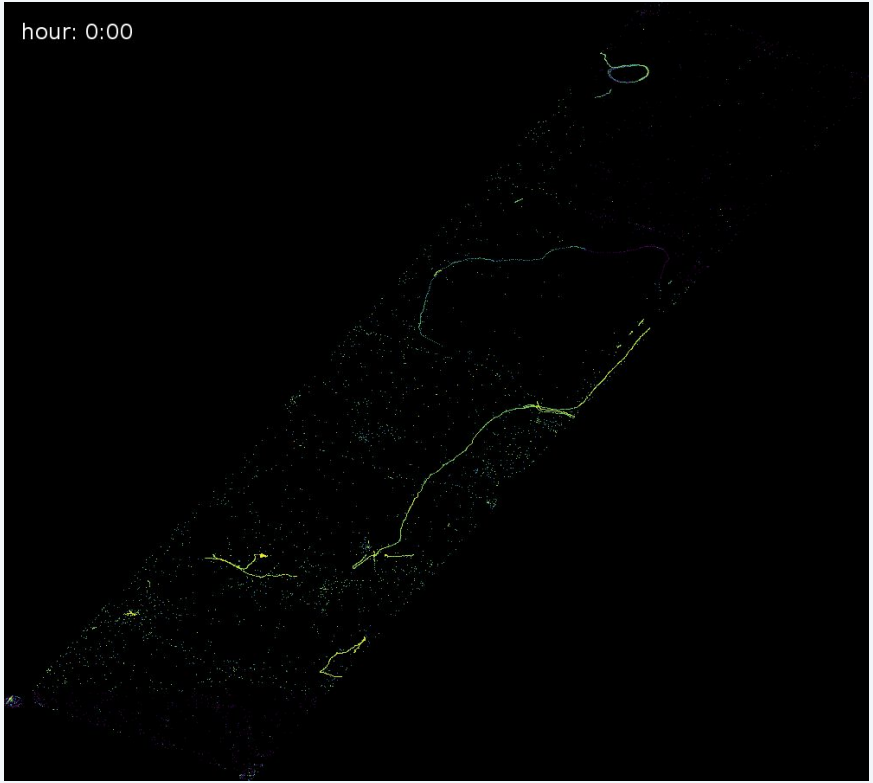
Taxi Trips to/from major airports around NYC

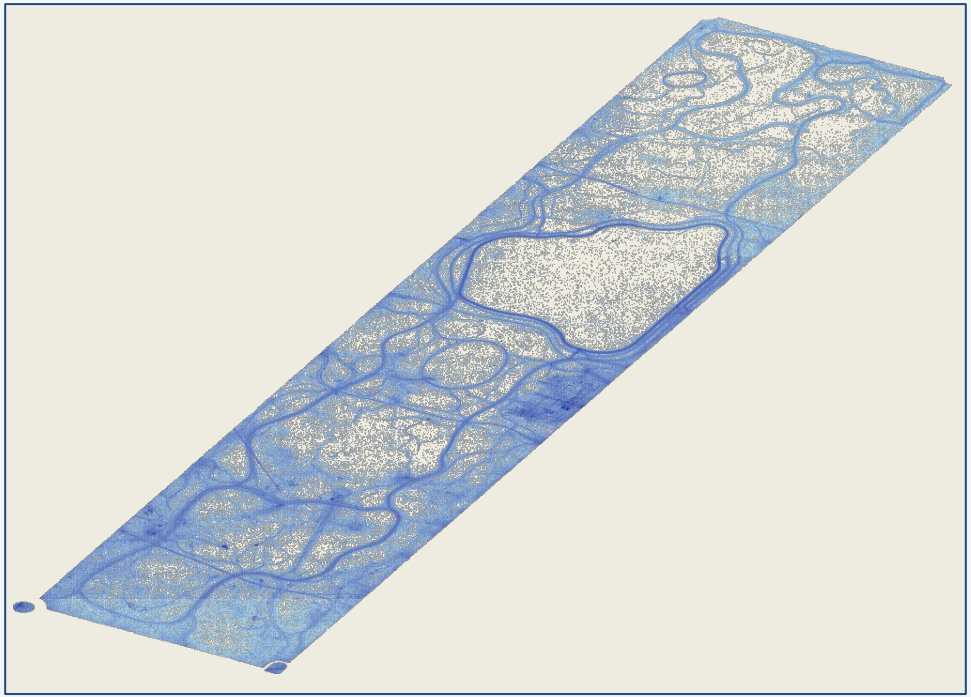




Mobility Data

`Lat/Lng/Time/Id`





Maps by Wenfei Xu

Data in the Spatial Structure

The geometries and their positions relative to one another provides additional data

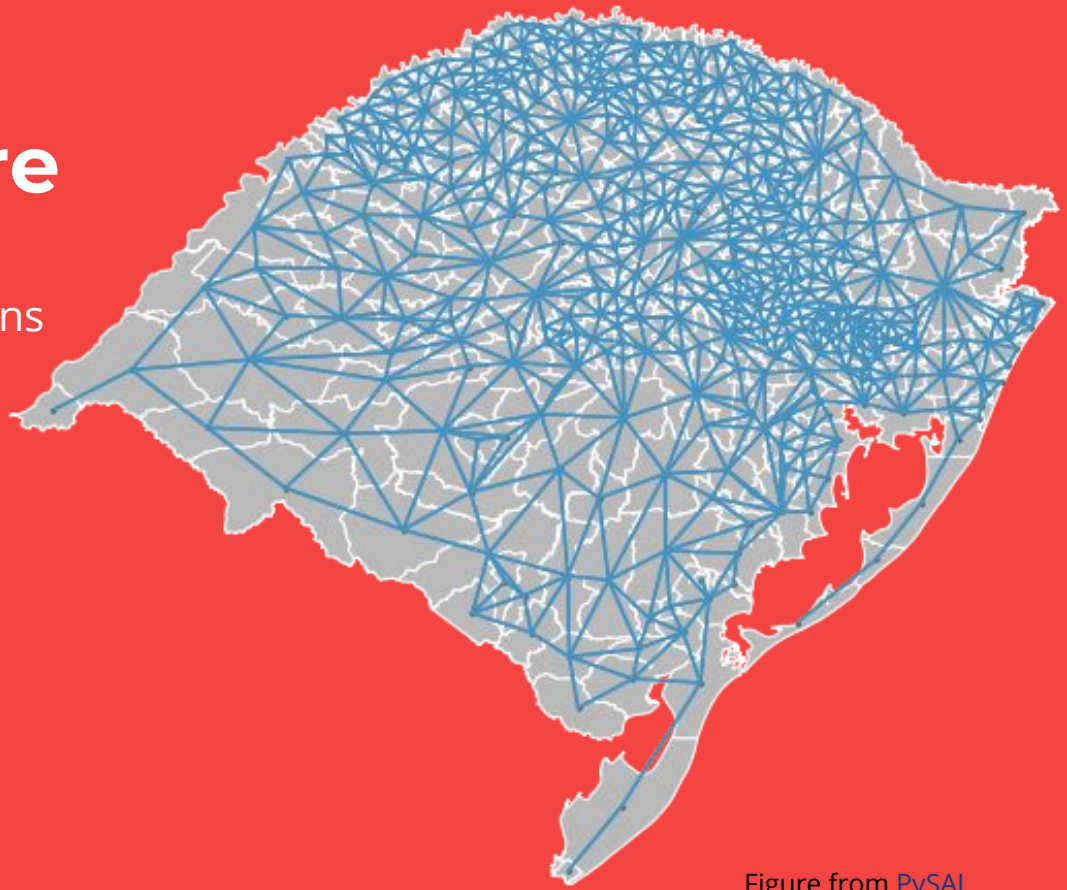
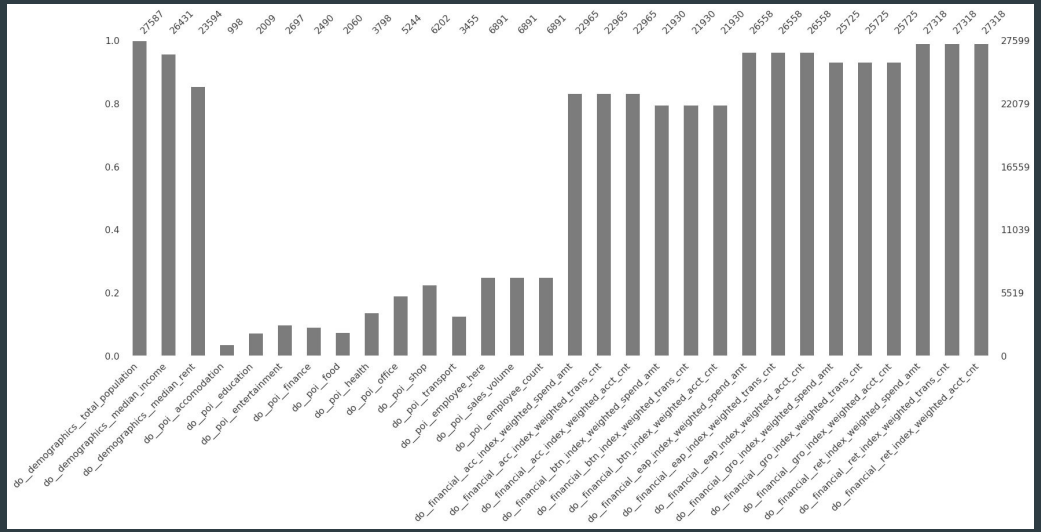


Figure from PySAL

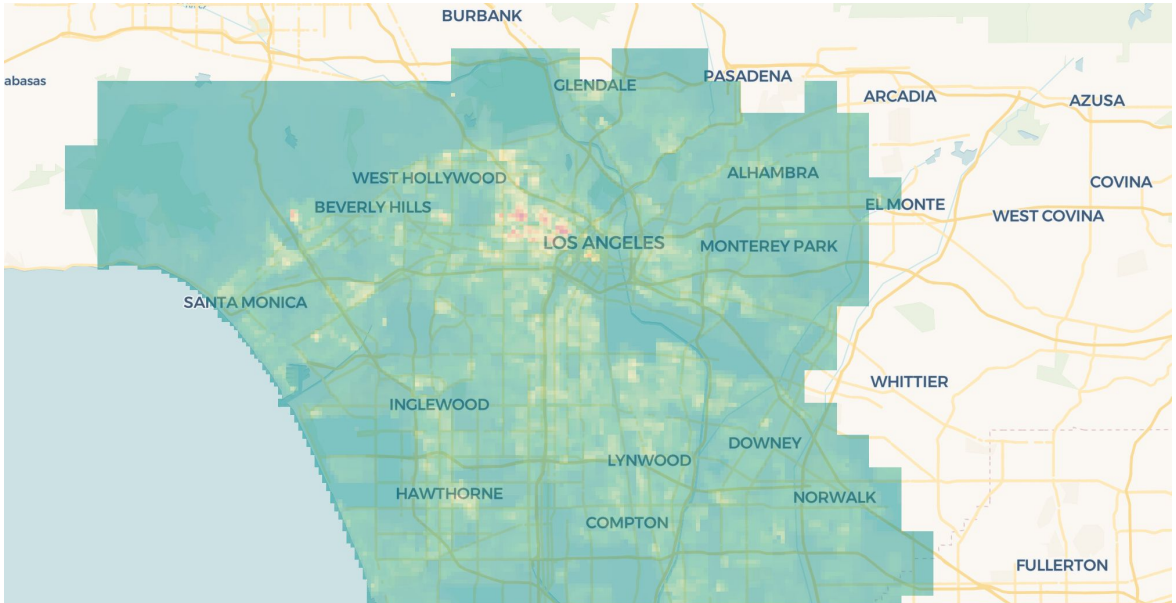
Working with missing data

Missing data is a common problem

- Missing because of geographic anonymization
- Lack of measurements at locations
- Data is messy



Given WeWork locations in NYC, show me potentially successful locations in LA



What data do we have?

- **WeWork locations**
  - 58 spaces in NYC
  - 21 in LA
- **Demographic** from the census
- **Financial data** from Mastercard's Retail Location Index
- **Points of Interest** (POI) for venues with similar characteristics (accomodation, eduction, food, entertainment, etc.)

Given WeWork locations in NYC, show me potentially successful locations in LA

**Compute distances in parameter space,
rank potential sites by similarity**

$$d\left(\mathbf{Y}_{origin}(i), \mathbf{Y}_{target}(f)\right) = \sqrt{\sum_{j}\left(Y_{origin}(i)_j - Y_{target}(f)_j\right)^2}$$

But...

My data has different variances, scales

Comes from many sources, has different scales, etc.

Some of my data is correlated

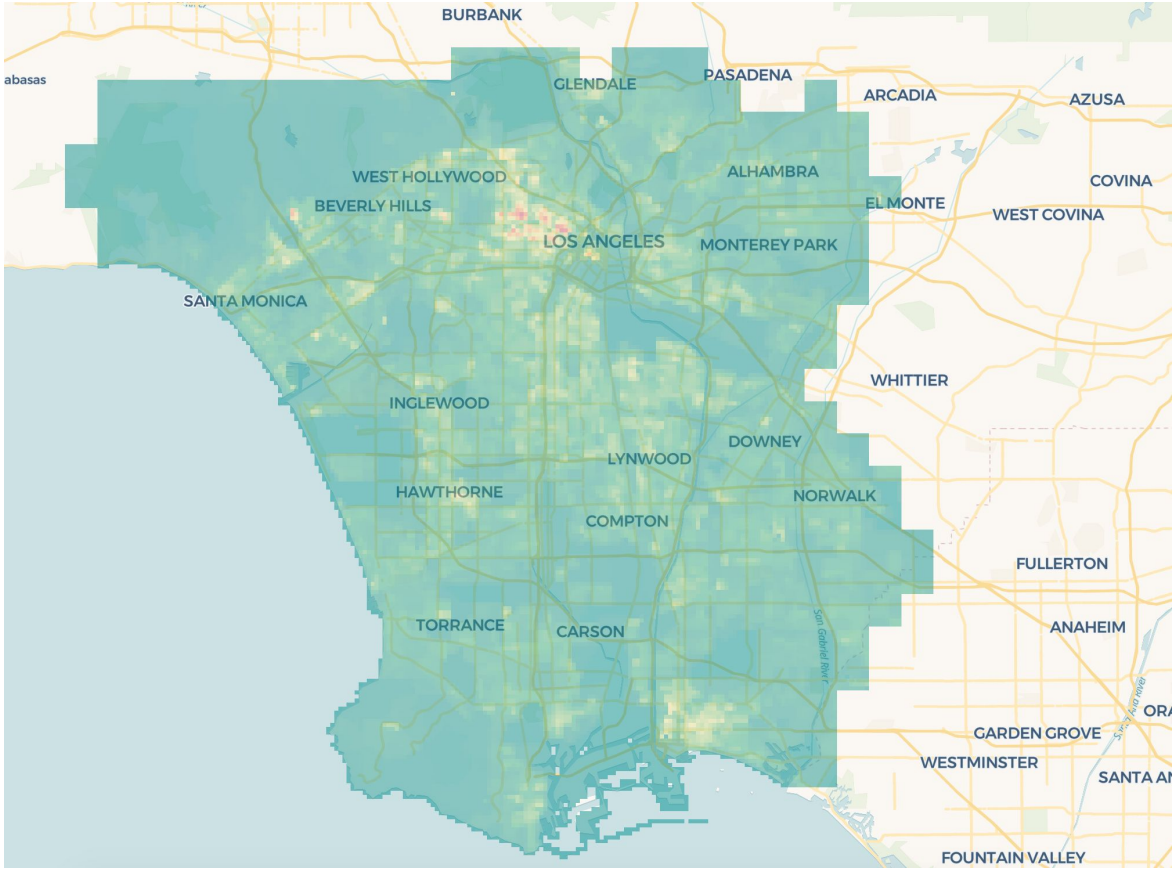
We should remove the redundancy due to correlation

I have missing values

Not all geographies have values, so we need to fill them in or remove those locations (not ideal)

Common Grid

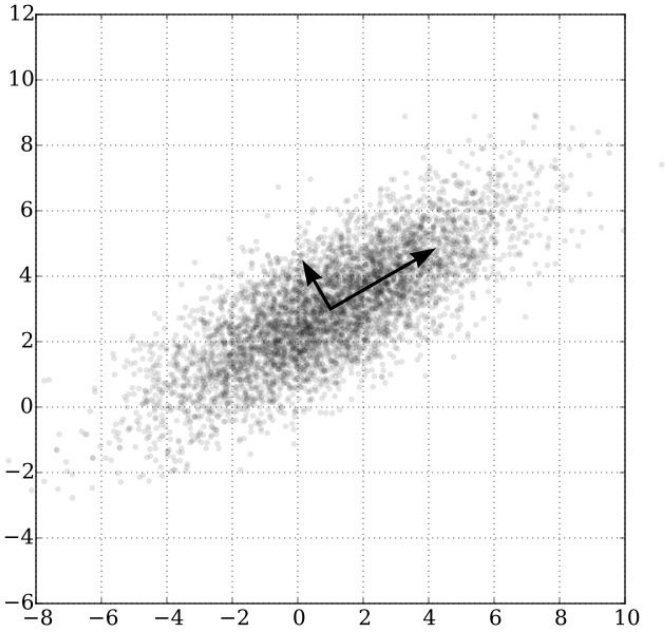
Use Quad Tree to hierarchically divide space, choose zoom level appropriate for aggregation



Principal Component Analysis (PCA)

Transform data to set of orthogonal axes (eigen decomposition)

- ★ Transformed features, including correlated ones, are linearly independent
- ★ Drop axes that explain least variance in data up to a threshold
- ✘ Doesn't work if data is missing

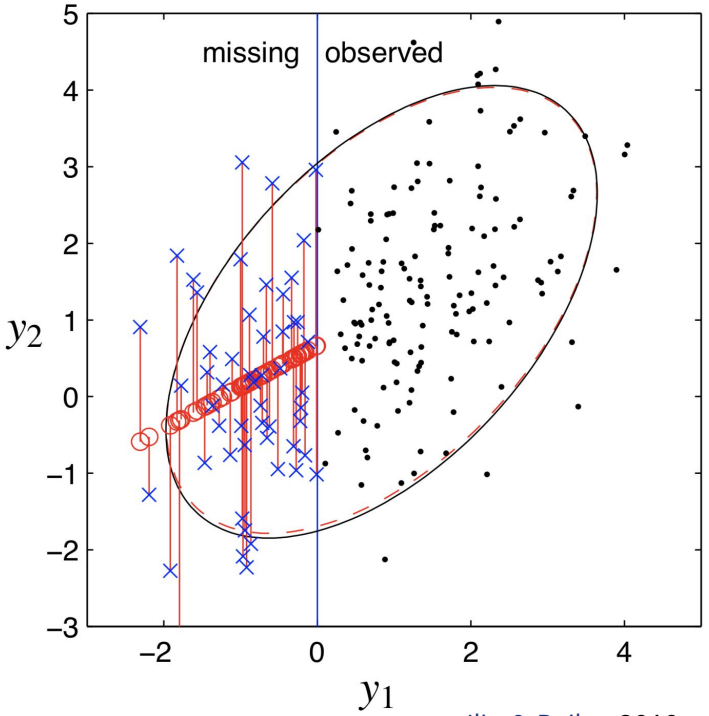


Probabilistic PCA

PCA doesn't work if we have missing data.

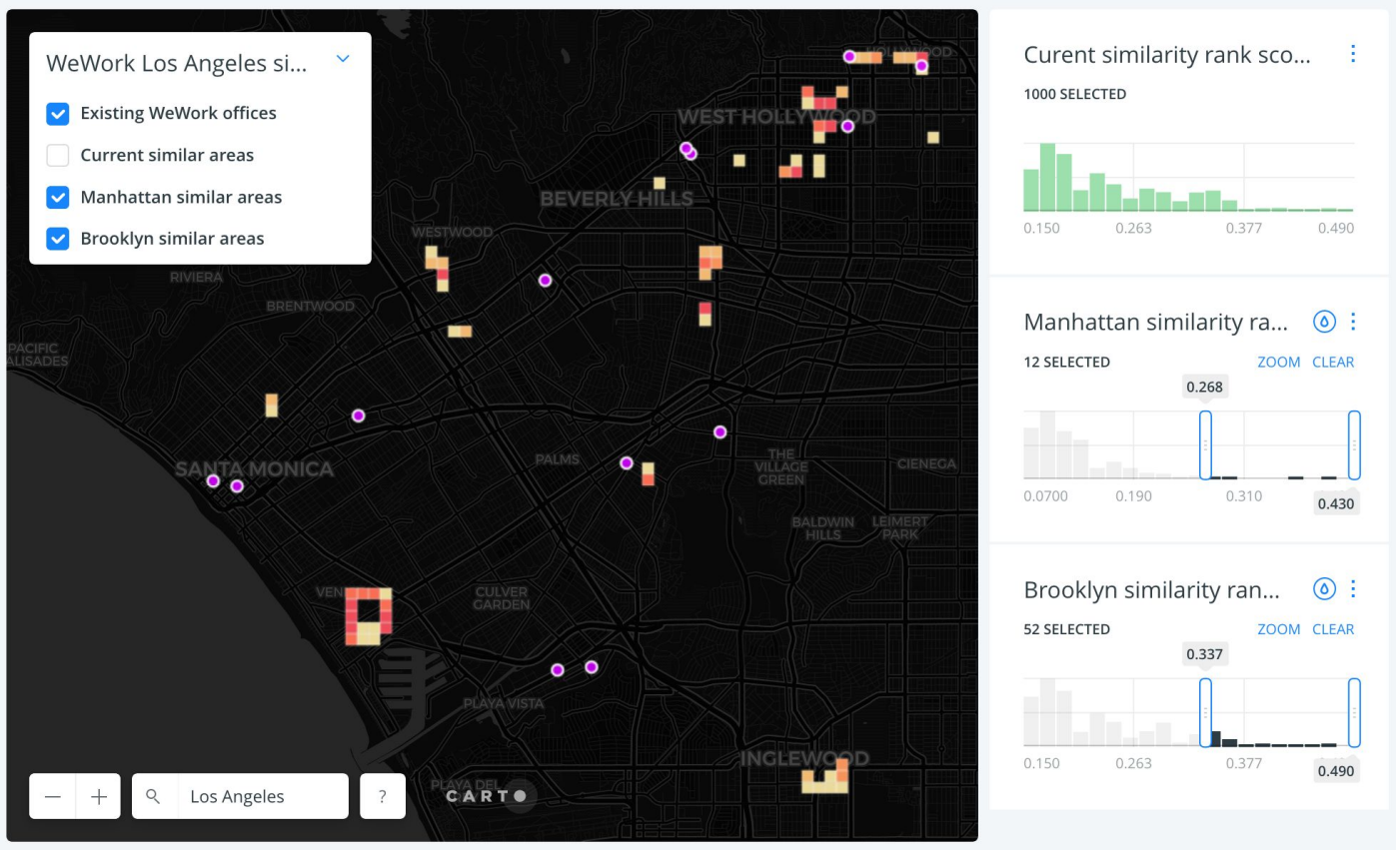
Common imputation falls short for more sizeable amounts of missing data

PPCA reconstructs the distribution of the data using the known data as a sample



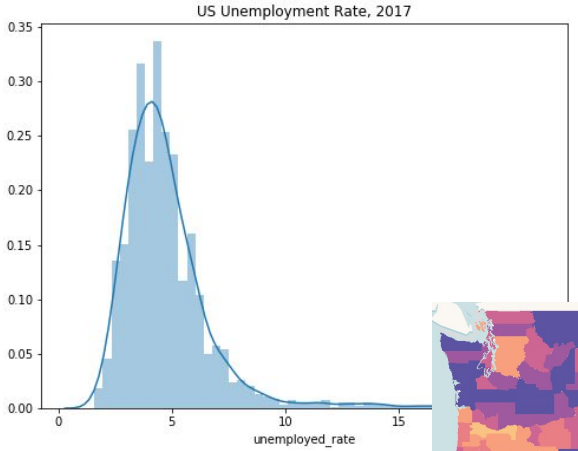
Ilin & Raiko, 2010

Results

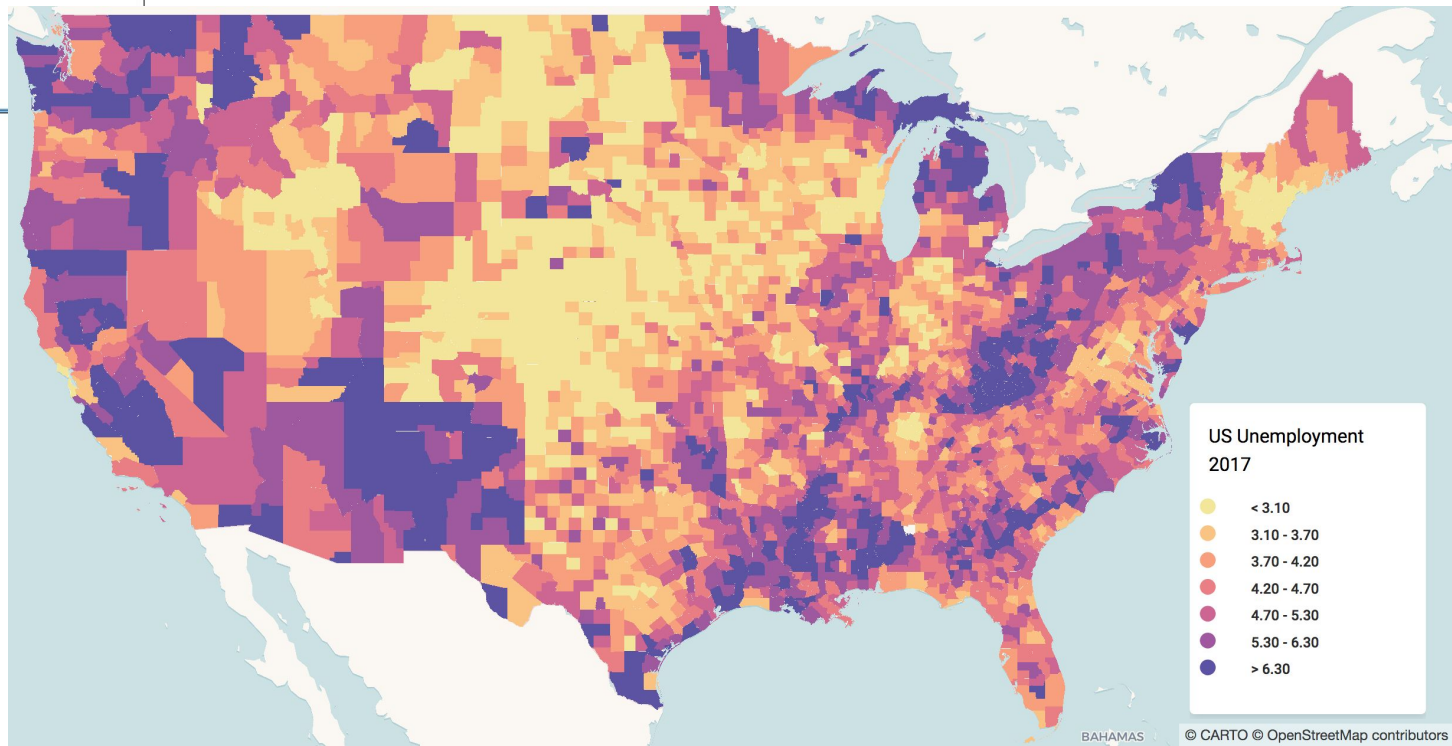


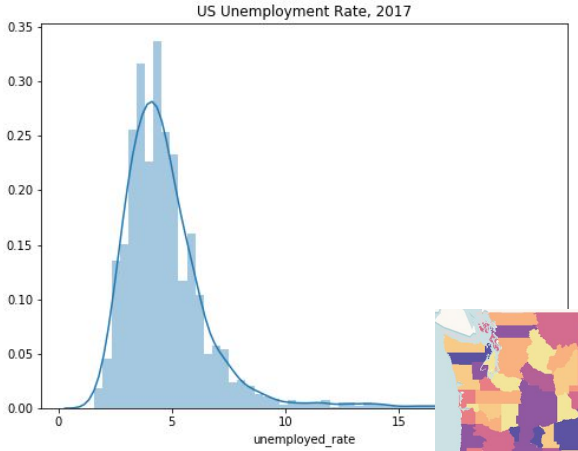
Analysis by Giulia Carella

Structure of Spatial Data

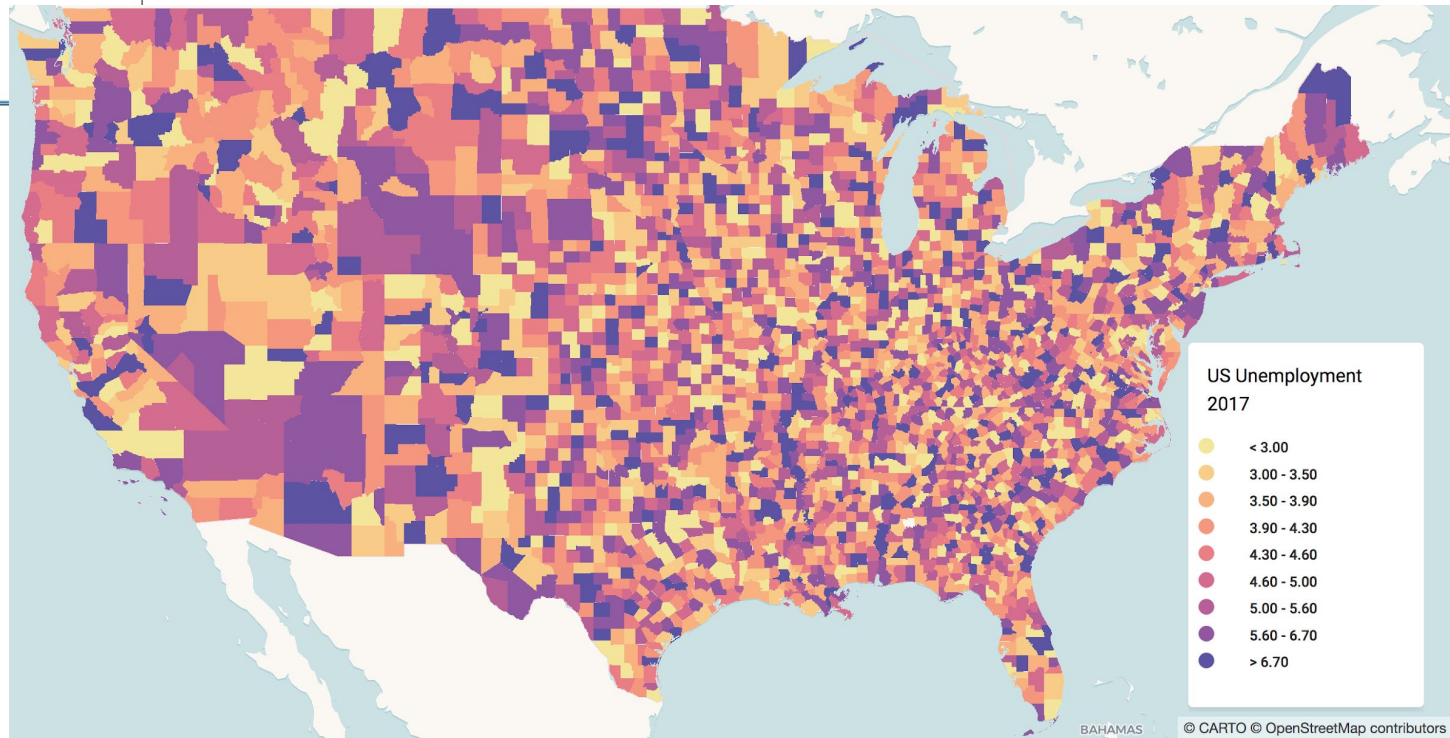

US Unemployment Rate, 2017
0.35
0.30
0.25
0.20
0.15
0.10
0.05
0.00
0
5
10
15
unemployed_rate

2019


US Unemployment
2017
< 3.10
3.10 - 3.70
3.70 - 4.20
4.20 - 4.70
4.70 - 5.30
5.30 - 6.30
> 6.30
BAHAMAS
© CARTO © OpenStreetMap contributors


US Unemployment Rate, 2017
0.35
0.30
0.25
0.20
0.15
0.10
0.05
0.00
0
5
10
15
unemployed_rate

2019

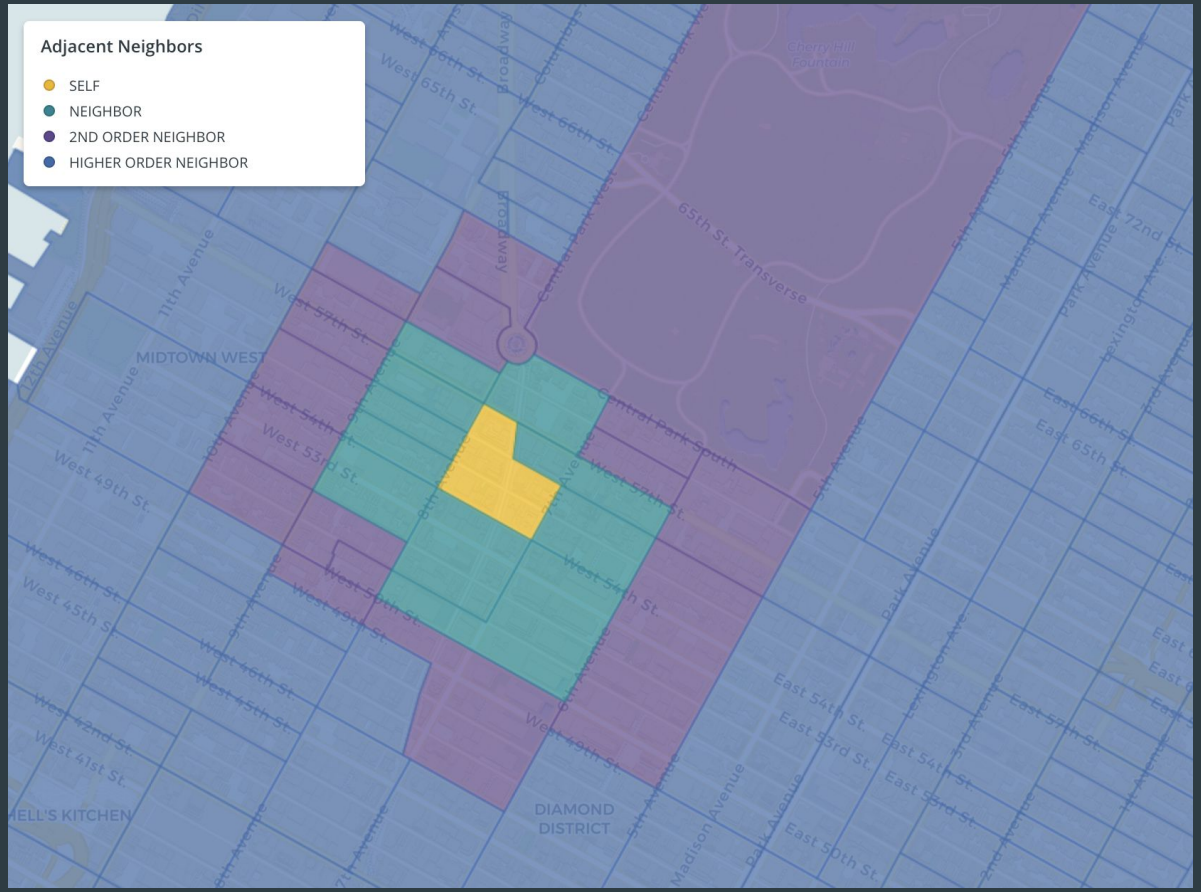

US Unemployment
2017
< 3.00
3.00 - 3.50
3.50 - 3.90
3.90 - 4.30
4.30 - 4.60
4.60 - 5.00
5.00 - 5.60
5.60 - 6.70
> 6.70
BAHAMAS
© CARTO © OpenStreetMap contributors

Spatial Weights

- Contiguity
- Distance
- kNN

Spatial Weights

Weights are built by 'neighbors', which is problem-dependent in how they are defined



Spatial Autocorrelation

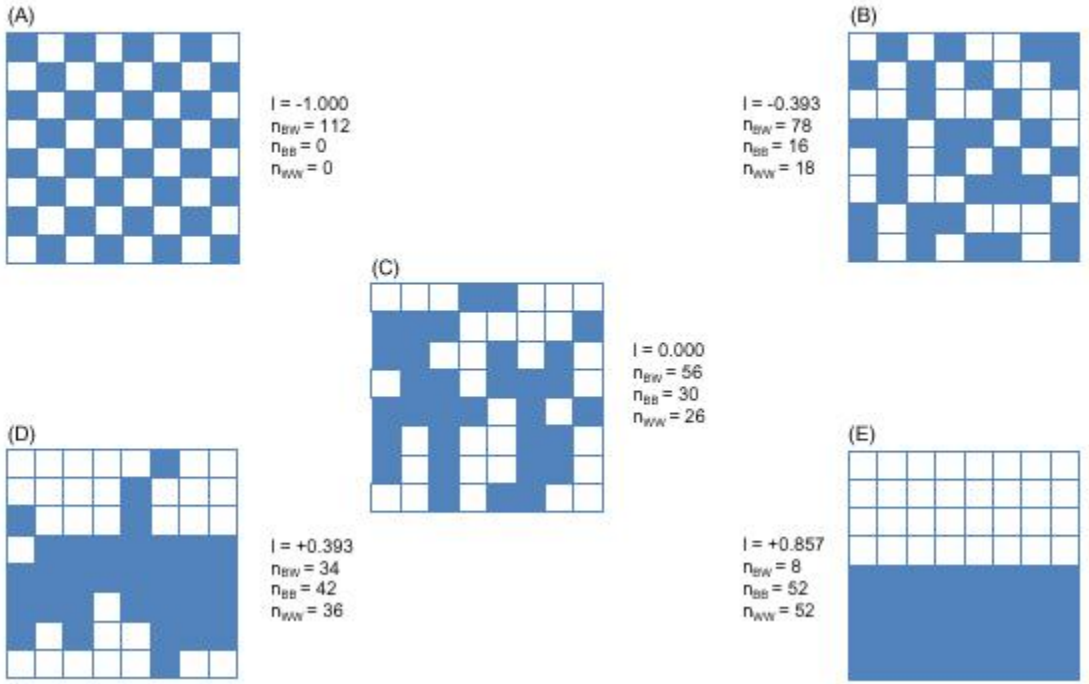
Moran's I statistic

Basic statistic for calculating the amount of:

- Clustering
- Outliers

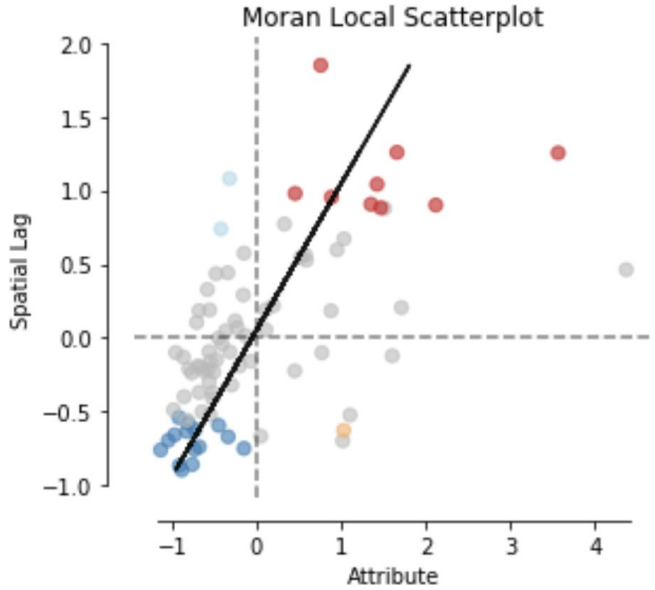
$$I = \frac{N}{W} \frac{\sum_i \sum_j w_{ij}(x_i - \bar{x})(x_j - \bar{x})}{\sum_i (x_i - \bar{x})^2}$$

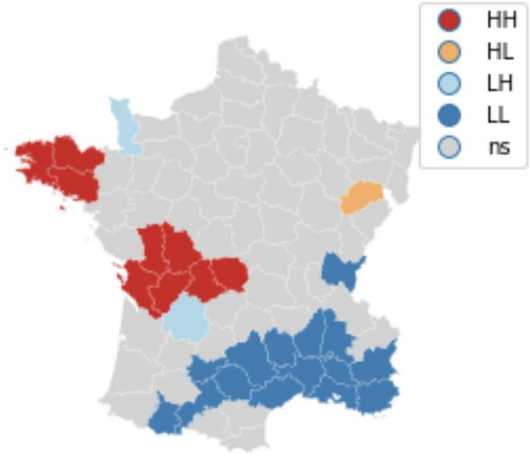
Spatial Autocorrelation



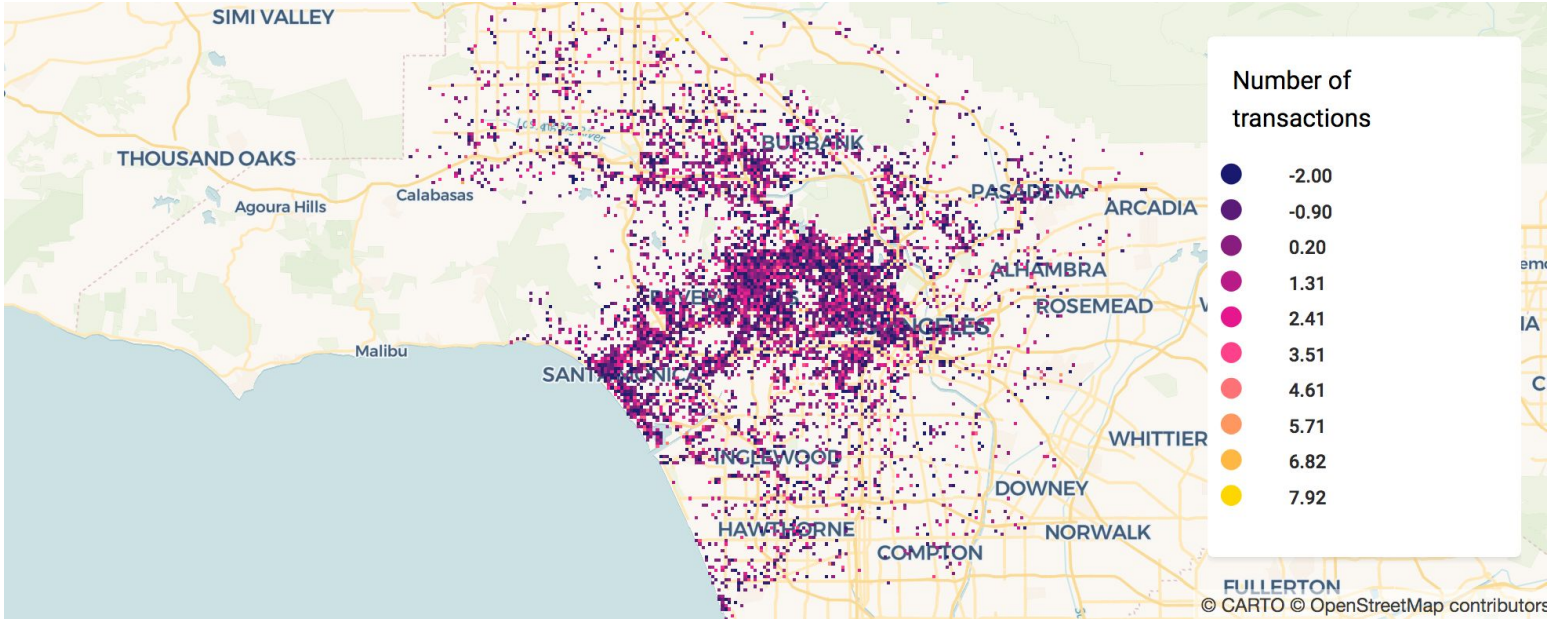
Spatial Autocorrelation (Local)

How a geometry compare to its neighbors





Measuring spatial residuals



CARTO

2019

Thanks!

Andy Eschbacher
Data Scientist
@MrEPhysics

Chapter one

Use this layout only if you have a lot of things to say - be mindful

We strongly suggest you to only use this slide if you absolutely need to.

The Earth was small, light blue, and so touchingly alone, our home that must be defended like a holy relic. **The Earth was absolutely round.** I believe I never knew what the word round meant until I saw Earth from space.

When I orbited the Earth in a spaceship, I saw for the first time how beautiful our planet is. Mankind, let us preserve and increase this beauty, and not destroy it!

Use this layout only if you have a lot of things to say - be mindful

We strongly suggest you to only use this slide if you absolutely need to.

The Earth was small, light blue, and so touchingly alone, our home that must be defended like a holy relic. **The Earth was absolutely round.** I believe I never knew what the word round meant until I saw Earth from space.

When I orbited the Earth in a spaceship, I saw for the first time how beautiful our planet is. Mankind, let us preserve and increase this beauty, and not destroy it!

What’s better than a list?

- The Earth was small, light blue, and so touchingly alone, our home that must be defended like a holy relic.
- The Earth was absolutely round.
- I believe I never knew what the word round meant until I saw Earth from space.

What’s better than a list?

- The Earth was small, light blue, and so touchingly alone, our home that must be defended like a holy relic.
- The Earth was absolutely round.
- I believe I never knew what the word round meant until I saw Earth from space.

What’s better than a list?

- The Earth was small, light blue, and so touchingly alone, our home that must be defended like a holy relic.
- The Earth was absolutely round.
- I believe I never knew what the word round meant until I saw Earth from space.

A numbered list!

1. The Earth was small, light blue, and so touchingly alone, our home that must be defended like a holy relic.

2. The Earth was absolutely round.

3. I believe I never knew what the word round meant until I saw Earth from space.

A numbered list!

1. The Earth was small, light blue, and so touchingly alone, our home that must be defended like a holy relic.

2. The Earth was absolutely round.

3. I believe I never knew what the word round meant until I saw Earth from space.

Hypnosis Myth Reality

Column 1. The Earth was small, light blue, and so touchingly alone, our home that must be defended like a holy relic.

The Earth was absolutely round. I believe I never knew what the word round meant until I saw Earth from space.

Column 2. A self-service business user application for spatial analysis and visualization.

Builder’s drag and drop analytics empower business analysts to optimize operations and quickly deploy location applications.

Hypnosis Myth Reality

Column 1. The Earth was small, light blue, and so touchingly alone, our home that must be defended like a holy relic.

The Earth was absolutely round. I believe I never knew what the word round meant until I saw Earth from space.

Column 2. A self-service business user application for spatial analysis and visualization.

Builder’s drag and drop analytics empower business analysts to optimize operations and quickly deploy location applications.

Peace On Earth A Wonderful Wish But No Way

Custom basemaps

Customized raster and vector maps that support worldwide coverage.

Geocoding

Multiple geocoding and permanent storage options

Routing

Global turn-by-turn directions for driving, biking, and walking.

Data Observatory

Added-value services like Demographics and Segmentation APIs

Peace On Earth A Wonderful Wish But No Way

Custom basemaps

Customized raster and vector maps that support worldwide coverage.

Geocoding

Multiple geocoding and permanent storage options

Routing

Global turn-by-turn directions for driving, biking, and walking.

Data Observatory

Added-value services like Demographics and Segmentation APIs

Hypnosis Myth Reality

The Earth was small, light blue, and so touchingly alone, our home that must be defended like a holy relic.

The Earth was absolutely round.

Custom basemaps

Customized raster and vector maps that support worldwide coverage.

Geocoding

Multiple geocoding and permanent storage options

Routing

Global turn-by-turn directions for driving, biking, and walking.

Data Observatory

Added-value services like Demographics and Segmentation APIs

Hypnosis Myth Reality

The Earth was small, light blue, and so touchingly alone, our home that must be defended like a holy relic.

The Earth was absolutely round.

Custom basemaps

Customized raster and vector maps that support worldwide coverage.

Geocoding

Multiple geocoding and permanent storage options

Routing

Global turn-by-turn directions for driving, biking, and walking.

Data Observatory

Added-value services like Demographics and Segmentation APIs

“Lorem ipsum dolor sit amet, consectetur adipiscing elit, sed do eiusmod tempor incididunt ut labore et dolore magna aliqua ipsum dolor sit amet.”

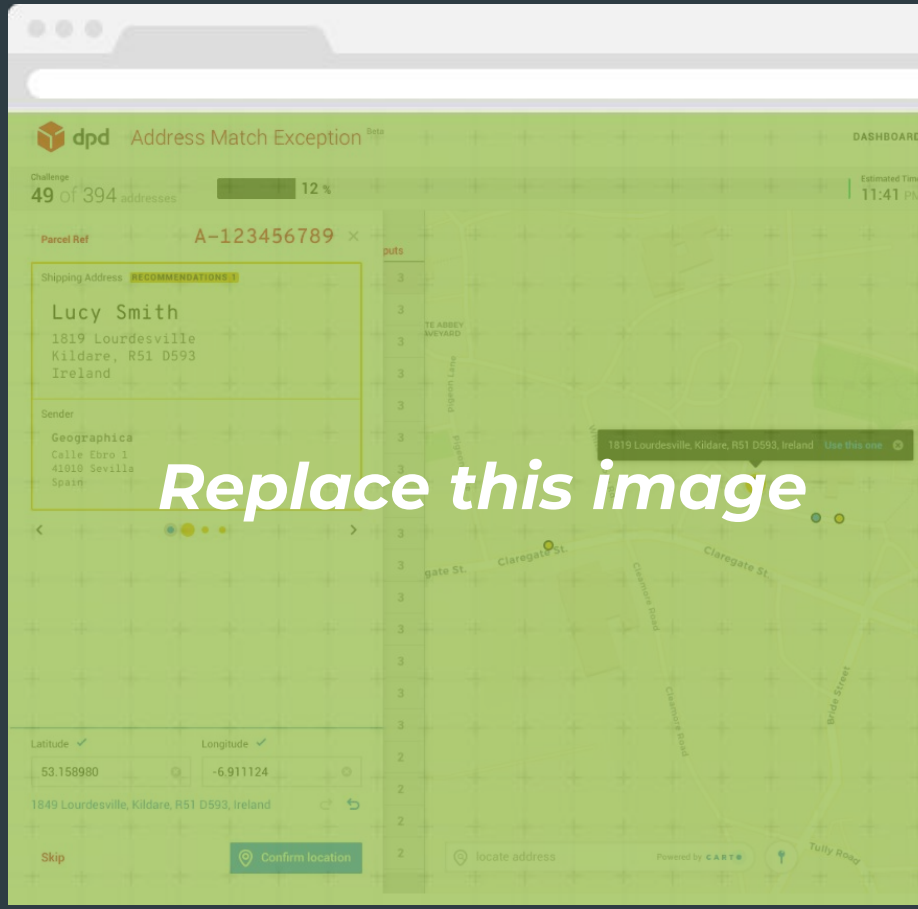
“Lorem ipsum dolor sit amet, consectetur adipiscing elit, sed do eiusmod tempor incididunt ut labore et dolore magna aliqua ipsum dolor sit amet.”

“Lorem ipsum dolor sit amet, consectetur adipiscing elit, sed do eiusmod tempor incididunt ut labore et dolore magna aliqua ipsum dolor sit amet.”

Logistics Optimization

Optimization of assignment of parcels and routes to drivers and their vehicles from their depots.

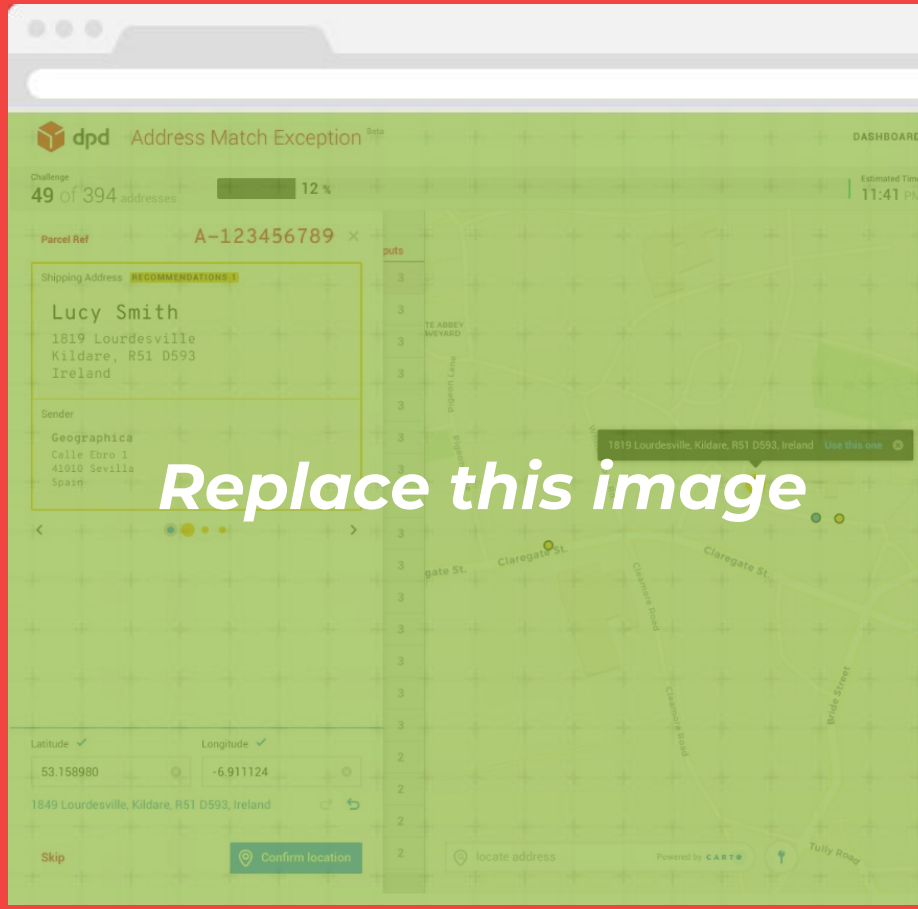
Every vehicle not needed saves them $150K/year



Logistics Optimization

Optimization of assignment of parcels and routes to drivers and their vehicles from their depots.

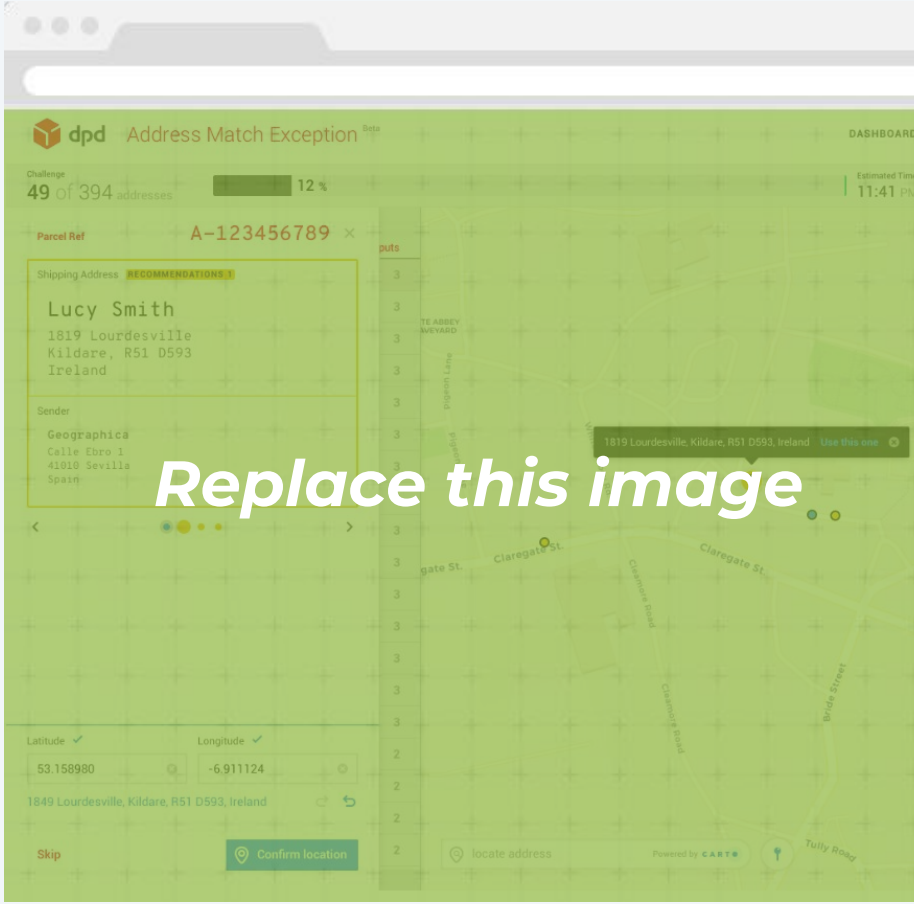
Every vehicle not needed saves them $150K/year



Logistics Optimization

Optimization of assignment of parcels and routes to drivers and their vehicles from their depots.

Every vehicle not needed saves them $150K/year



“Lorem ipsum dolor sit amet, consectetur adipiscing elit, sed do eiusmod tempor incididunt ut”

“Lorem ipsum dolor sit amet, consectetur adipiscing elit, sed do eiusmod tempor incididunt ut”

“Lorem ipsum dolor sit amet, consectetur adipiscing elit, sed do eiusmod tempor incididunt ut”

Escape the GIS Handcuffs

Text + 1 image. To replace the placeholder, just select the image and click on the “Replace image” option.

If you need to use an image smaller than the placeholder; simply click on ‘Insert Image’ and delete the placeholder once inserted.

Escape the GIS Handcuffs

Text + 1 image. To replace the placeholder, just select the image and click on the “Replace image” option.

If you need to use an image smaller than the placeholder; simply click on ‘Insert Image’ and delete the placeholder once inserted.

Escape the GIS Handcuffs

Text + 2 images. To replace the placeholders, just select each image and click on the “Replace image” option.

If you need to use an image smaller than the placeholder; simply click on ‘Insert Image’ and delete the placeholder once inserted.

Escape the GIS Handcuffs

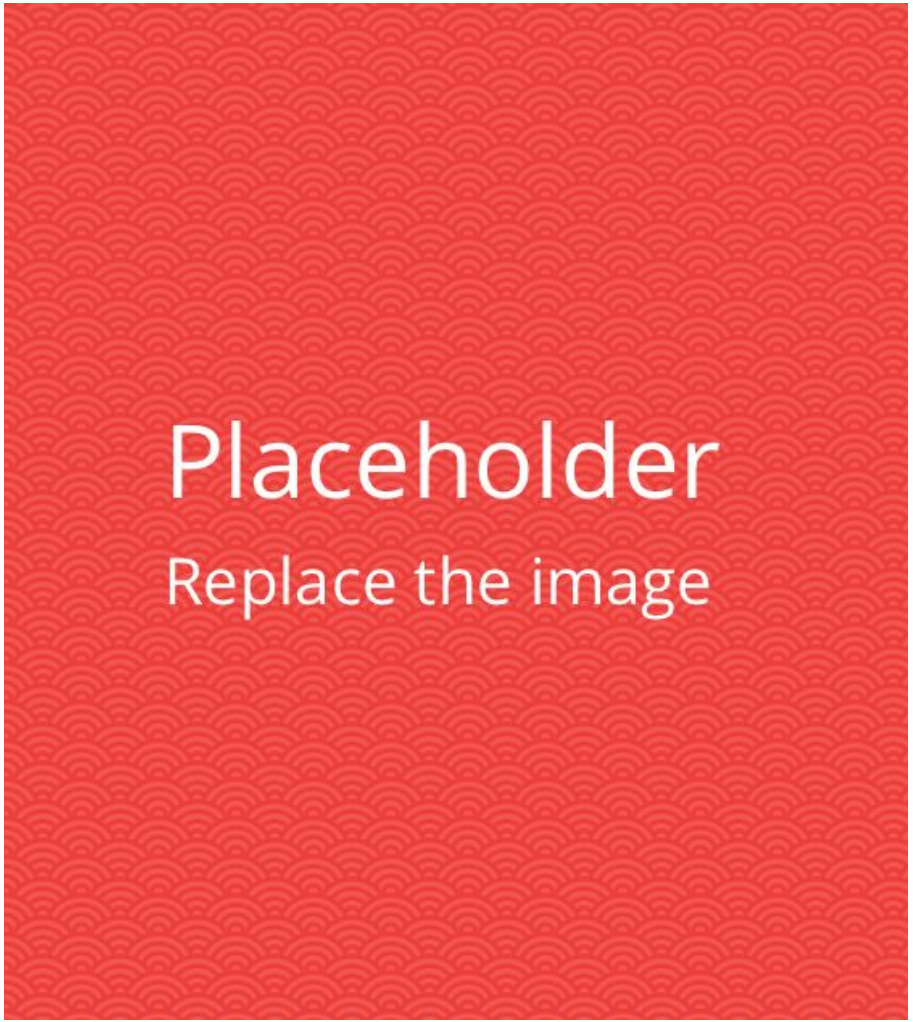
Text + 2 images. To replace the placeholders, just select each image and click on the “Replace image” option.

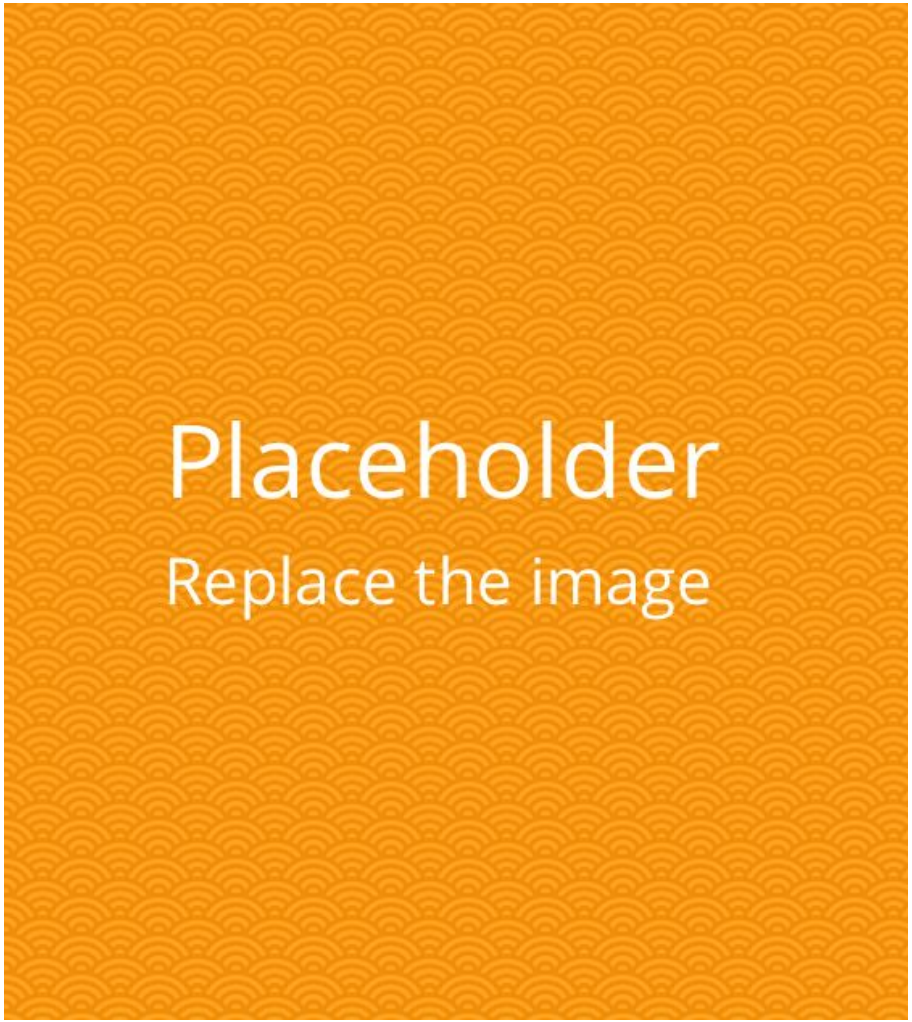
If you need to use an image smaller than the placeholder; simply click on ‘Insert Image’ and delete the placeholder once inserted.

Chapter one

Placeholder

Replace the image


Placeholder
Replace the image


Placeholder
Replace the image


Placeholder
Replace the image


Placeholder
Replace the image


Placeholder
Replace the image


Placeholder
Replace the image

Placeholder

Replace the image

Placeholder

Replace the image

Placeholder

Replace the image

Peace On Earth A Wonderful Wish But No Way

Remember	To	name	all	columns	ok
Item	Item	Item	Item	Item	Item
Item	Item	Item	Item	Item	Item
Item	Item	Item	Item	Item	Item
Item	Item	Item	Item	Item	Item
Item	Item	Item	Item	Item	Item

Counting Your Chicken Before They Hatch

Placeholder

Replace the image


Placeholder
Replace the image

The Emerald Buddha

To replace the placeholder, click on the “Replace image” option. If you need to use an image smaller than the placeholder; simply click on ‘Insert Image’ and delete the placeholder once inserted.

Bienvenid@s

Pauline Becker
Demo Developer Intern

Madrid — April 24

Rhoda Obrien
Systems

Brooklyn — May 1

Henrietta Soto
Research & Data

Brooklyn — May 1

Bienvenid@s

Rhoda Obrien
Systems

Brooklyn — May 1

Henrietta Soto
Research & Data

Brooklyn — May 1

Bienvenid@

Rhoda Obrien

Systems

Brooklyn — May 1

Kudos

Pauline Becker
Demo Developer Intern

Rhoda Obrien
Systems

Henrietta Soto
Research & Data

CARTOVersaries

Pauline Becker

Demo Developer Intern

2 Yrs — Nov 23

Rhoda Obrien

Systems

1 Year — Nov 29

Henrietta Soto

Research & Data

2 Years — Nov 30

Hasta luego

Pauline Becker

Demo Developer Intern

Last Day — July 3

Rhoda Obrien

Systems

Last Day — July 12

Henrietta Soto

Research & Data

Last Day — July 13

We’re hiring!

DON’T FORGET! You could get 1,500 if you refer someone we hire!

Sales

Account Exec (Financial Sector)
Washington, DC

Sales Development Rep
Washington, DC

Product, Eng. & Tech

Front-end Engine & Builder
Madrid

Back-end Engine & Builder
Madrid

We’re hiring!

DON’T FORGET! You could get 1,500 if you refer someone we hire!

Sales

Account Exec (Financial Sector)
Washington, DC

Sales Development Rep
Washington, DC

Product, Eng. & Tech

Front-end Engine & Builder
Madrid

We’re hiring!

DON’T FORGET! You could get 1,500 if you refer someone we hire!

Sales

Account Exec (Financial Sector)
Washington, DC

Product, Eng. & Tech

Front-end Engine & Builder
Madrid

We’re hiring!

DON’T FORGET! You could get 1,500 if you refer someone we hire!

Sales

Account Exec (Financial Sector)
Washington, DC

Escape the GIS Handcuffs

Text + code. Just replace the code with the one you need

Builder's drag and drop analytics empower business analysts to optimize operations and quickly deploy location applications.

```
#nyc_block_group {
  polygon-fill: #FFFFB2;
  polygon-opacity: 0.8;
  line-color: #FFF;
  line-width: 0.5;
  line-opacity: 1;

    [ commute_60min_more <= 1] {
      polygon-fill: #B10026;
    }
    [ commute_60min_more <=
0.367032967032967] {
      polygon-fill: #E31A1C;
    }
    [ commute_60min_more <=
0.283171521035599] {
      polygon-fill: #FC4E2A;
    }
}
```

```
#nyc_block_group {
  polygon-fill: #FFFFB2;
  polygon-opacity: 0.8;
  line-color: #FFF;
  line-width: 0.5;
  line-opacity: 1;

    [ commute_60min_more <= 1] {
      polygon-fill: #B10026;
    }
    [ commute_60min_more <=
0.367032967032967] {
      polygon-fill: #E31A1C;
    }
    [ commute_60min_more <=
0.283171521035599] {
      polygon-fill: #FC4E2A;
    }
}
```

Escape the GIS Handcuffs

Text + code. Just replace the code with the one you need.

Builder's drag and drop analytics empower business analysts to optimize operations and quickly deploy location applications.

```
#nyc_block_group {
  polygon-fill: #FFFFB2;
  polygon-opacity: 0.8;
  line-color: #FFF;
  line-width: 0.5;
  line-opacity: 1;

    [ commute_60min_more <= 1] {
      polygon-fill: #B10026;
    }
    [ commute_60min_more <= 0.3670329670329670] {
      polygon-fill: #E31A1C;
    }
    [ commute_60min_more <= 0.2831715210355990] {
      polygon-fill: #FC4E2A;
    }
}
```

If you really need to use this

The Earth was small, light blue, and so touchingly alone, our home that must be defended like a holy relic.

A self-service business user application for spatial analysis and visualization. Builder's drag and drop analytics empower business analysts to optimize operations and quickly deploy location applications.

Builder's drag and drop analytics empower business analysts to optimize operations and quickly deploy location applications.

Just select the image and click on the "Replace image" option.

Free yourself from reliance on GIS specialists and put the power of location intelligence directly in your hands.

If you really need to use this

The Earth was small, light blue, and so touchingly alone, our home that must be defended like a holy relic.

A self-service business user application for spatial analysis and visualization. Builder's drag and drop analytics empower business analysts to optimize operations and quickly deploy location applications.

Builder's drag and drop analytics empower business analysts to optimize operations and quickly deploy location applications.

Just select the image and click on the "Replace image" option.

Free yourself from reliance on GIS specialists and put the power of location intelligence directly in your hands.

If you really need to use this

The Earth was small, light blue, and so touchingly alone, our home that must be defended like a holy relic.

A self-service business user application for spatial analysis and visualization. Builder's drag and drop analytics empower business analysts to optimize operations and quickly deploy location applications.

Builder's drag and drop analytics empower business analysts to optimize operations and quickly deploy location applications.

Just select the image and click on the "Replace image" option.

Free yourself from reliance on GIS specialists and put the power of location intelligence directly in your hands.

Questions?

Questions?

Questions?

CARTO

2019

Thanks !

CARTO

2019

Thanks !