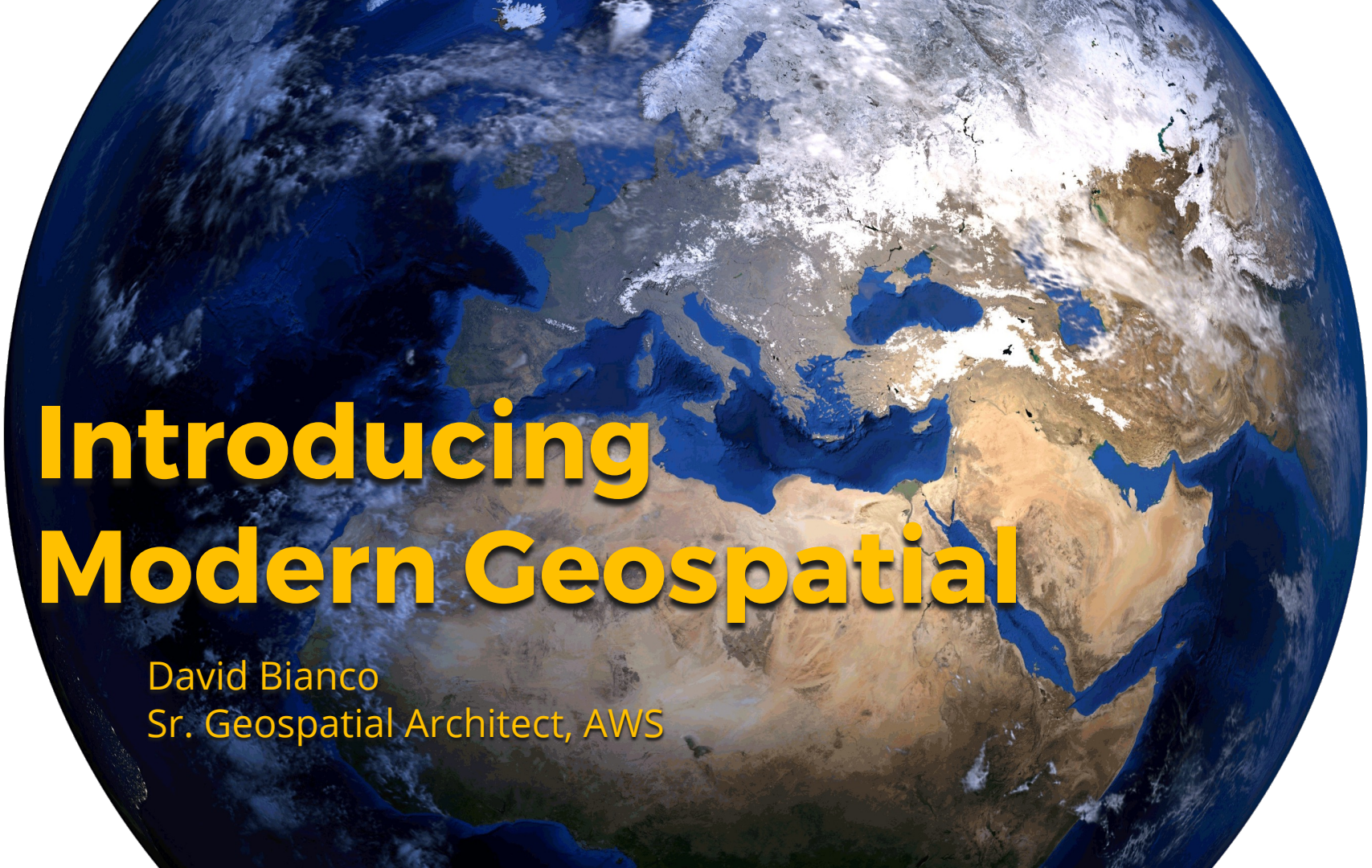
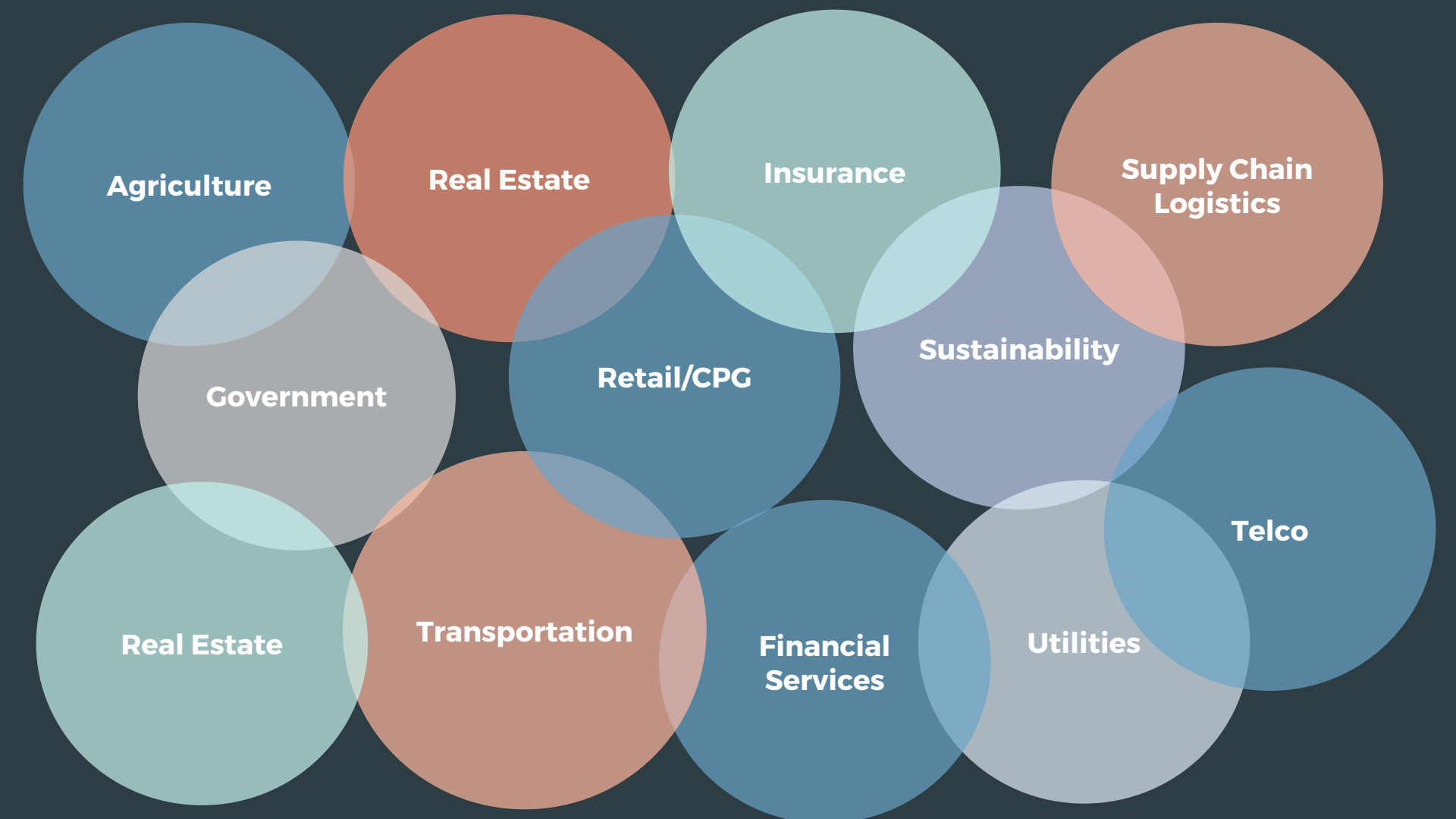


Introducing Modern Geospatial

David Bianco
Sr. Geospatial Architect, AWS


Agriculture
Real Estate
Insurance
Supply Chain Logistics
Government
Retail/CPG
Sustainability
Telco
Real Estate
Transportation
Financial Services
Utilities

A brief history of GIS

John Snow Cholera Map

British Physician John Snow began mapping cholera outbreak locations, finding that they were associated with contaminated water, not air.

The birth of Spatial Analysis.

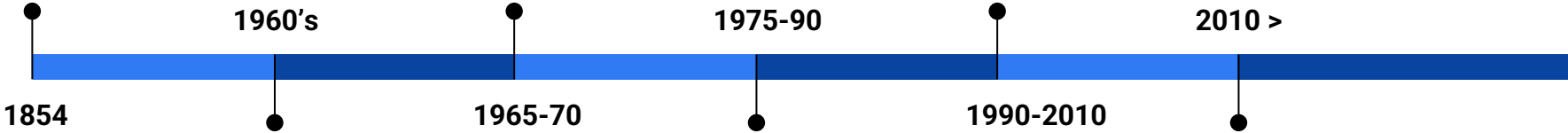
Pioneering GIS

Advancements in technology including:

- Map graphics as outputs using line printers
- Advances in data storage with mainframe computers
- Recording coordinates as input data

User Proliferation

- Cheaper more powerful compute
- Multiple software options
- The launch of new satellites and remote sensing technology
- Publicly available GPS revolutionises data capture



Sieve Mapping

The use of transparent layers on light tables to overlay and intersect different geographic data sources.

Many drawbacks incentivised the shift from paper to computer mapping.

The commercialisation of GIS

ESRI sprang from the Harvard Computer Graphics Laboratory and would go on to become the dominant player in the market.

Roger Tomlinson first used the term "Geographic Information System" in his publication in 1968 "A Geographic Information System for Regional Planning".

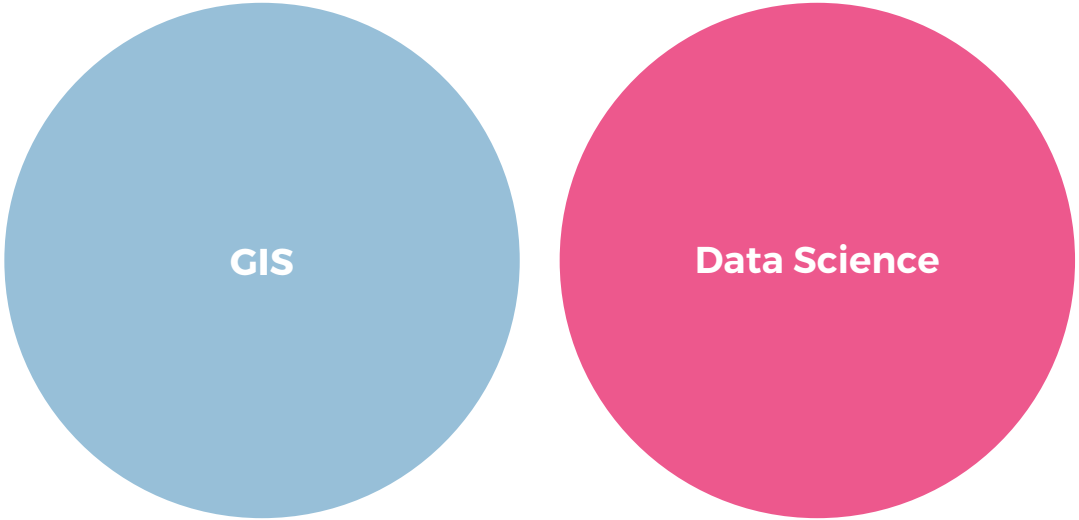
The open source revolution

- GIS data becomes ubiquitous and freely available
- Open Street Map
- Pioneering open source developments including QGIS, GRASS, GDAL, PostgreSQL & PostGIS
- Proliferation of Geospatial Python Libraries & Integration into mainstream tech stacks

Legacy GIS

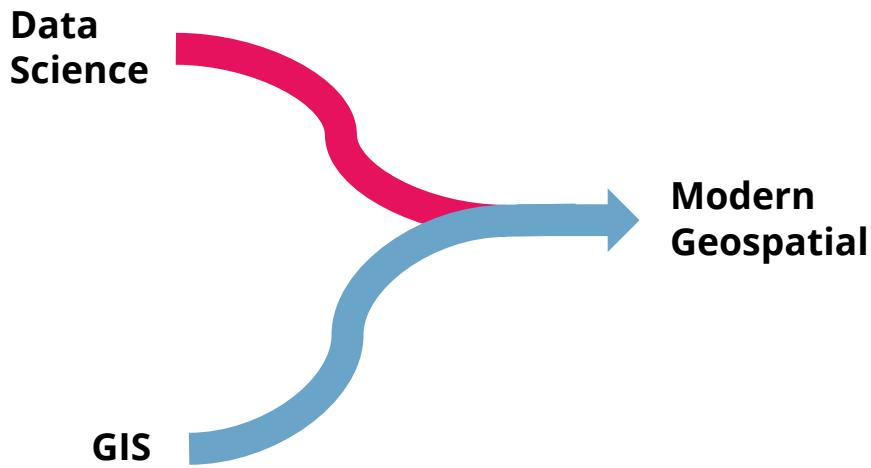
- Monopolistic market: ESRI
- Walled garden / Closed Source proprietary software
- Pseudo open file formats

- Siloed teams
- Outputs typically static & focussed around maps



Modern geospatial

- Standards based
- Open & Interoperable Tools
- Familiar programming languages (Python, SQL, R)
- Directly integrated with typical Data Science tech stacks
- Thriving open source ecosystem
- Cloud native & highly scalable for:
    - Massive data volumes
    - Different data types & velocities



Modern Geospatial is *Cloud-Native Data Formats*

Cloud-Optimized GeoTIFF (COG)

- Efficient imagery data access
- Legacy compatibility
- Cloud native, optimized for S3
- Bring compute to your data
- Rasterio python library
- https://cogeo.org/

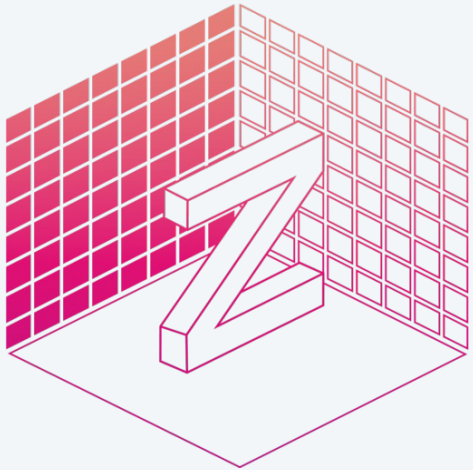
Modern Geospatial is *Cloud-Native Data Formats*

Cloud-Optimized Point Cloud (COPC)

- Manage LiDAR point cloud data
- LAZ file (LAS zip compressed)
- Cloud native, optimized for S3
- Friends with PDAL, Entwine, Greyhound
- https://copc.io/

Modern Geospatial is *Cloud-Native Data Formats*

Zarr



- Common format for climate analytics use cases
- N-dimensional array
- Cloud-native (S3) vs HDF, NetCDF
- https://zarr.dev/

Modern Geospatial is *Cloud-Native Data Formats*

GeoParquet

- A powerful column-oriented data format
- Built from the ground up as an alternative to the CSV
- Cross platform
- An incubating OGC Standard to add geospatial data types (point, line, polygon) to Parquet
- https://github.com/opengeospatial/geoparquet

Modern Geospatial is *Community*

OGC®
Making location count.

PANGEO

OSGeo
Your Open Source Compass

CLOUD-NATIVE GEOSPATIAL FOUNDATION

slack
The Spatial Community

ANALYSIS READY DATA

Modern Geospatial is
Analytics at Scale

Data Warehouse

- Analytics use cases
- SQL engine
- Serverless and Scalable options
- DuckDB – the new kid in town

Modern Geospatial is
Analytics at Scale

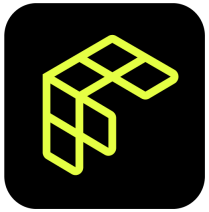
Notebooks

- Perfect for experiments, R&D
- Analytics and visualization
- Collaboration
- Self-contained

CARTO

Amazon SageMaker

databricks



Fused.io

Modern Geospatial is
Analytics at Scale

Parallel Compute

- Python
- Spark, Scala
- Big data
- Serverless functions

Modern Geospatial is *Serverless Web Maps*

- Raster and Vector map tiles
- TiTiler (raster)
- PMTiles (vector)
- Multiple front-end JS libraries
  - MapLibre
  - CARTO
  - Leaflet
  - OpenLayers

ProtoMaps

Modern Geospatial is *Open Data*

OVERTURE MAPS FOUNDATION

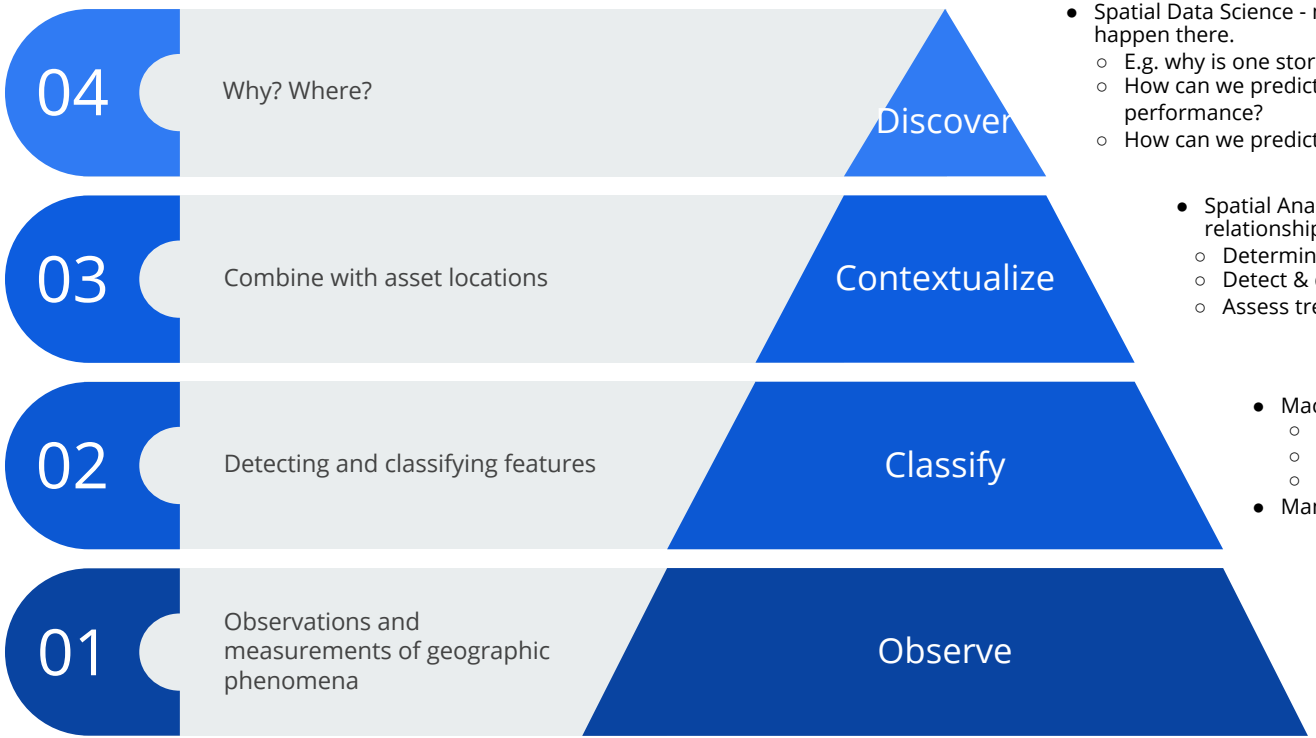
aws open data

CARTO Spatial Data Catalog

SOURCE COOPERATIVE

STAC SpatioTemporal Asset Catalog

The Spatial Pyramid



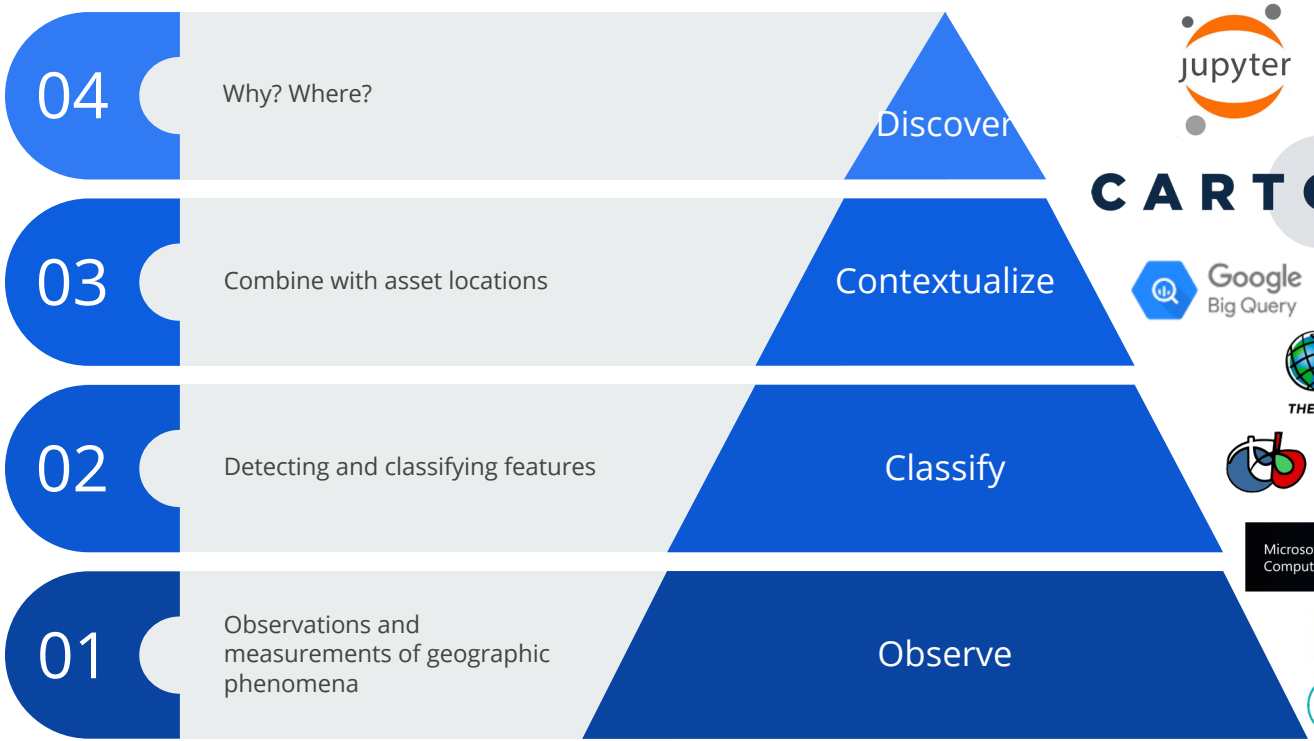
- Spatial Data Science - moving beyond where things happen to **why** they happen there.
  - E.g. why is one store in a portfolio outperforming another?
  - How can we predict new locations that will deliver similar performance?
  - How can we predict potential store revenue ahead of time?

- Spatial Analysis - studying the characteristics of places and the relationships between them:
  - Determine relationships
  - Detect & quantify patterns
  - Assess trends (spatio-temporal)

- Machine Learning Algorithms
  - Object Detection
  - Image Classification
  - Clustering & Aggregation
- Manual Digitisation - OpenStreetMap

- Raster Data (continuous):
  - Earth Observation (Satellites)
  - Visible, Hyperspectral, RADAR etc.
- Vector Data (discreet):
  - Government bodies & National statistics agencies
  - Commercial data providers e.g. Foot traffic, POI data.

The Spatial Pyramid - Toolkit Examples

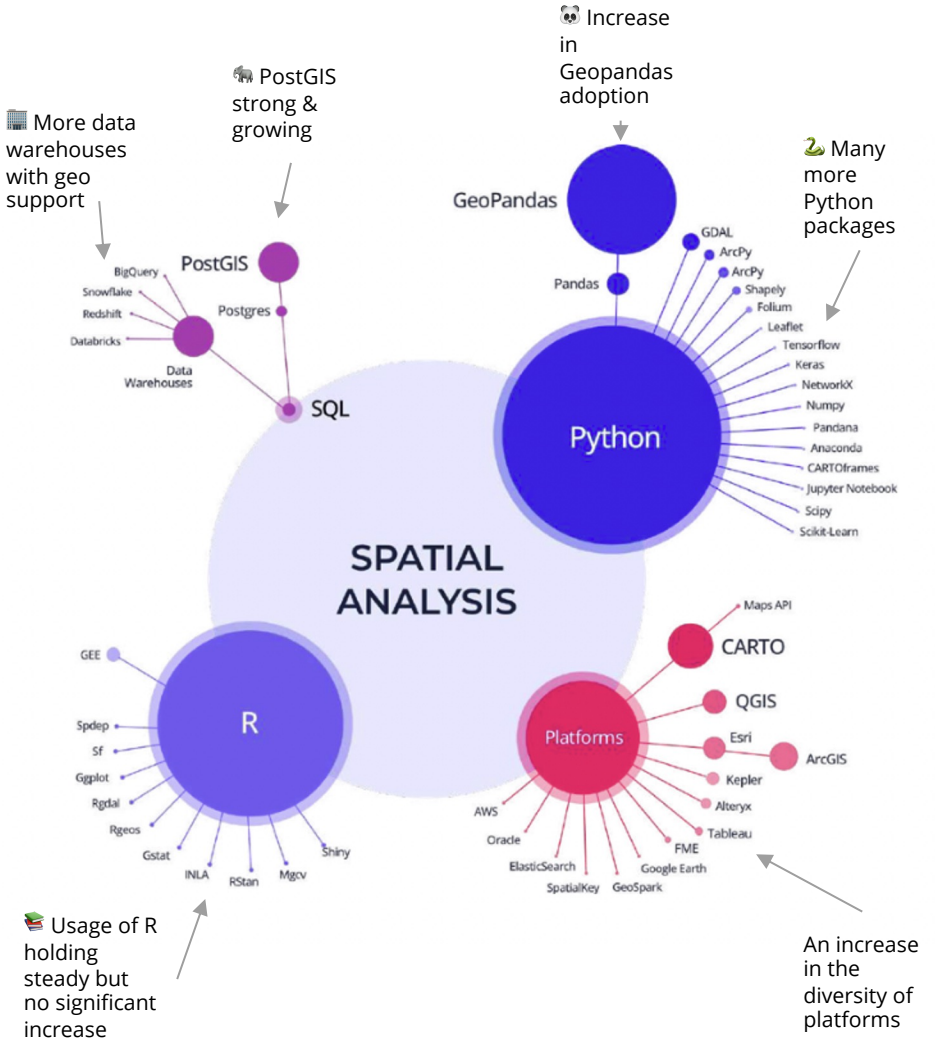


jupyter
NumPy
pandas
CARTO
PostGIS
Google Big Query
GeoPandas
GDAL
esri
THE SCIENCE OF WHERE
scikit learn
machine learning in Python
TensorFlow
Microsoft Planetary Computer
Google Earth Engine
Office for National Statistics
lifesight
here
planet.
NASA
CLIMATE TRACE.

The State of Spatial Data Science 2023 (vs. 2022)

💡86% of teams use Python or SQL as their primary programming language

From: The State of Spatial Data Science in Enterprise 2023 by CARTO & AWS.



The growth in Geo-Python packages

Applications are diverse:

From remote sensing image processing through to network analysis & routing.

Multiple packages are available for every stage of the workflow from object detection and image classification through to data visualisation.

2022 Geospatial Python Packages Stats

Percent increase is from 2021. Data provided by BigQuery Public Data - Python Package Index dataset.

geemap	1.2M	+843%
leafmap	88K	+355%
esda	336K	+217%
whitebox	298K	+190%
pydeck	10.7M	+127%
geopandas	32M	+108%
libpysal	591K	+105%
osmnx	954K	+92%
fiona	33M	+81%
folium	10.4M	+66%
shapely	92M	+63%
pyproj	56M	+58%
rasterio	10.9M	+46%
gdal	2.7M	+18%

Spatial SQL

Spatial SQL uses all the same elements and structure of normal SQL but allows you to work with another data type: a GEOMETRY or GEOGRAPHY.

GEOMETRIES can be classified in:

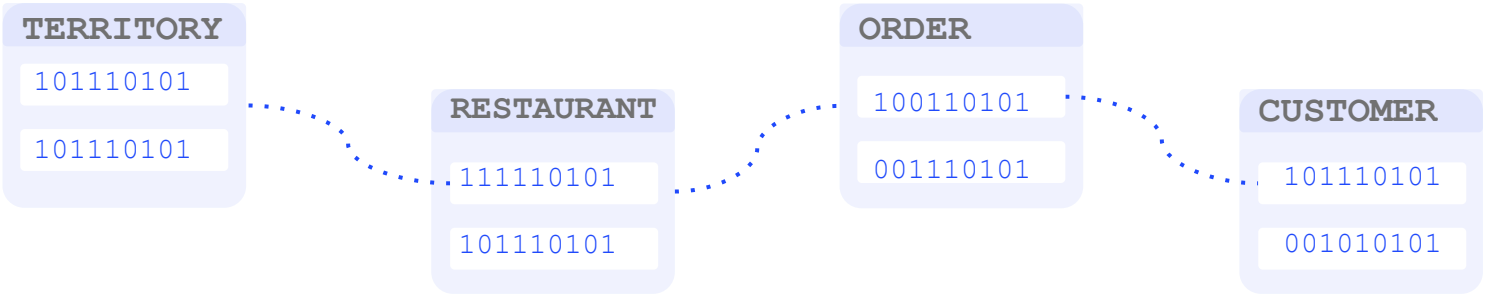
1. Points (pairs of coordinates)
2. Lines (sequence of coordinates)
3. Polygons (group of coordinates that form shapes)

With these new data types come a set of functions, commonly predicated with **ST** (such as ST_Intersects) which stands for spatial type.

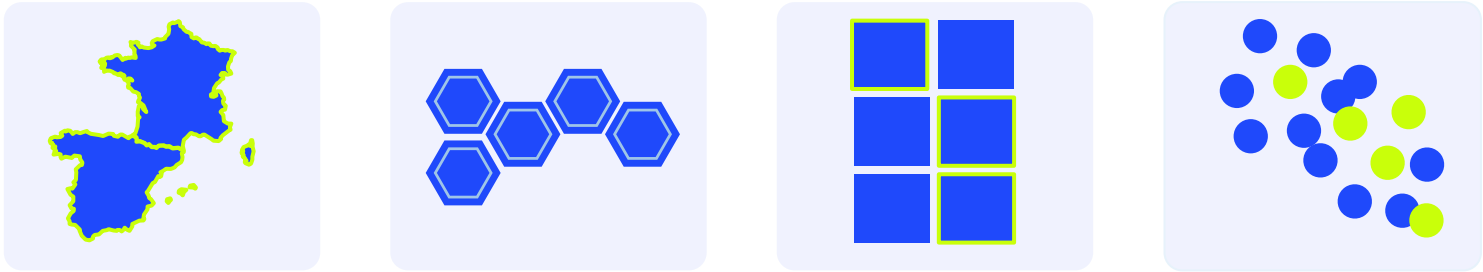
Spatial SQL - language of the Data Warehouse

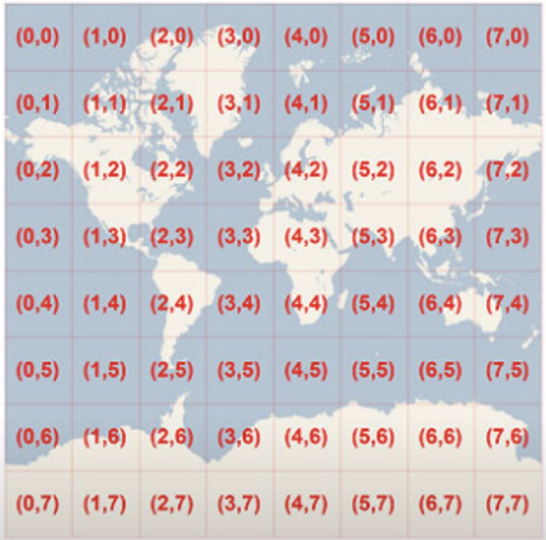
	Native ST Functions	CARTO's Analytics Toolbox Functions	Total
PostGIS	271	17	**288**
Google Big Query	64	123	**187**
amazon REDSHIFT	102	62	**164**
snowflake	66	82	**148**
databricks	-	71	**71**

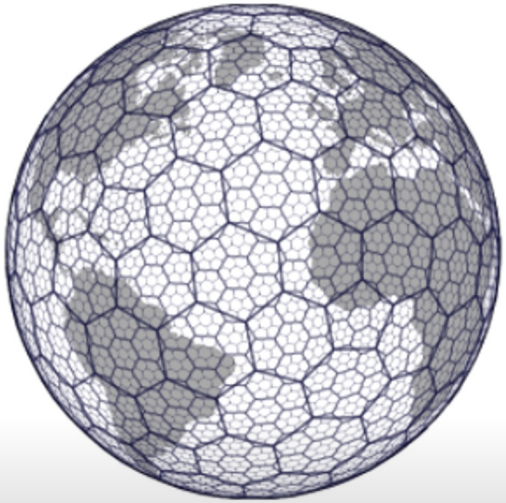
How do you model your data?

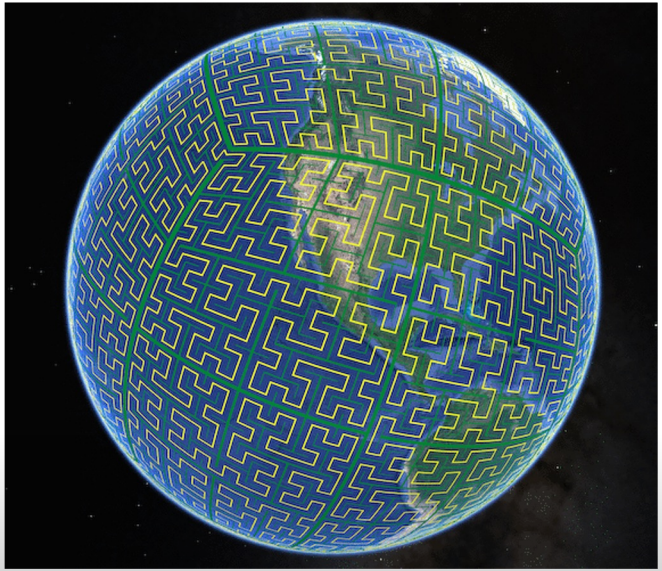


What geographic support you use matters a lot









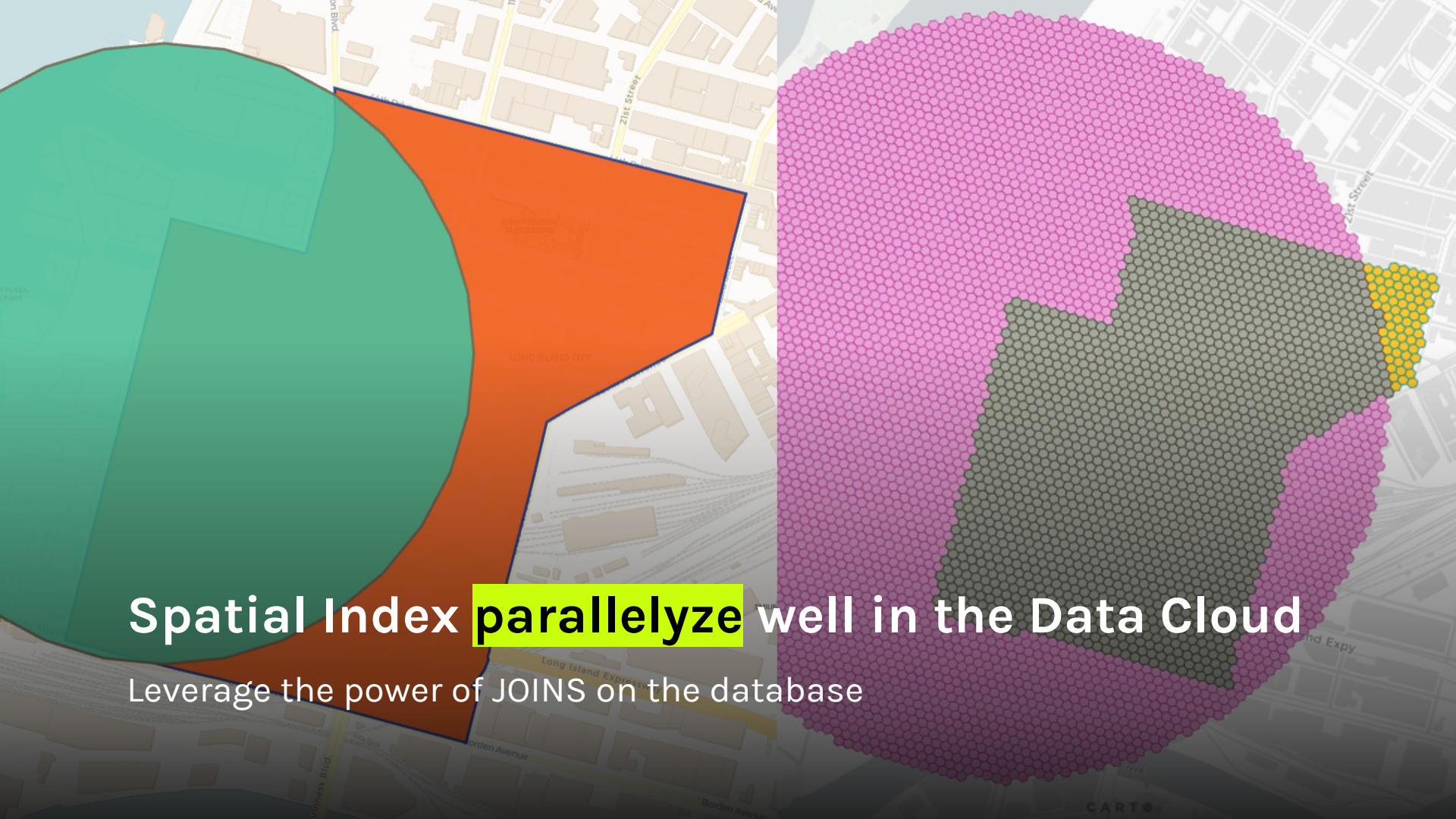
Quadkey (source)

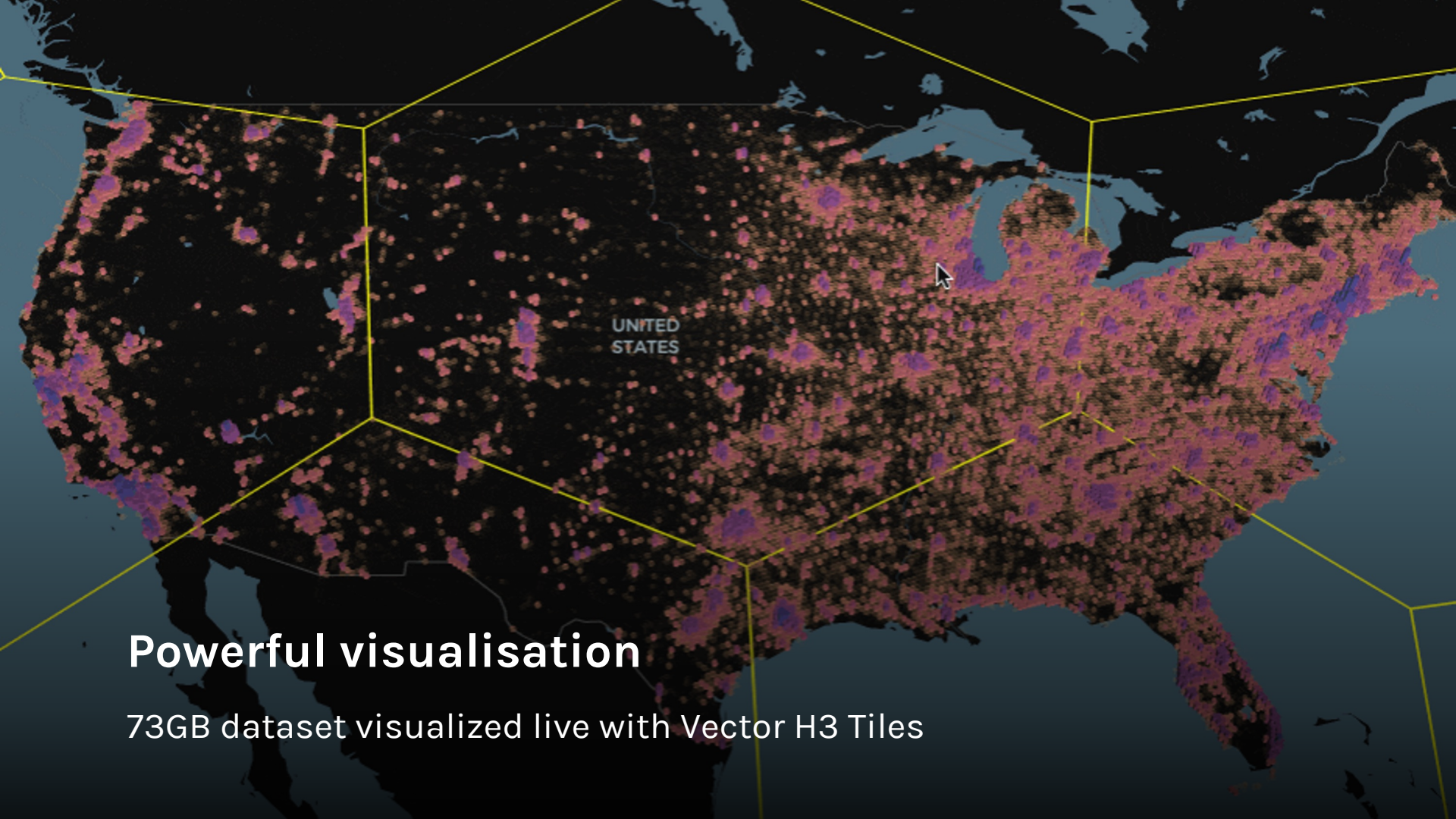
Uber's H3 (source)

S2 (source)

Geospatial Hierarchical Indexes

Different strategies to partition the space intro discrete grids


Spatial Index parallelyze well in the Data Cloud
Leverage the power of JOINS on the database



Powerful visualisation

73GB dataset visualized live with Vector H3 Tiles

Thank You.

David Bianco, AWS