

Geospatial Data

Lucía Santamaría

lucia.santamaria@ymail.com

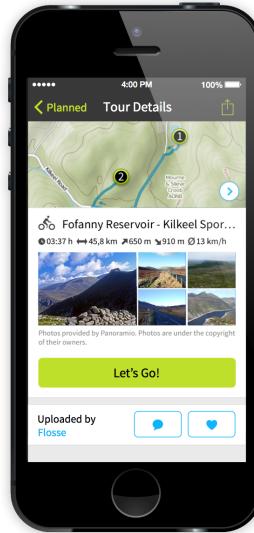


DATA SCIENCE RETREAT

TODAY

- Introduction: Geospatial data
- Planar geometries*
- Easy maps with leaflet/folium*
- Creating and loading a spatial database
- Visualizing data from postGIS to html*
- Exploring postGIS capabilities*

(*) Hands-on tutorials on Jupyter [<https://github.com/lucasantamaria/geodata>]



@lusantala

lucia.santamaria@ymail.com

Geospatial Data

Geospatial Data

is the data or information that identifies the **geographic location** of features and boundaries on Earth. Spatial data is usually stored as coordinates and topology, and is data that **can be mapped**. Spatial data is often accessed, manipulated or analyzed through Geographic Information Systems (**GIS**)

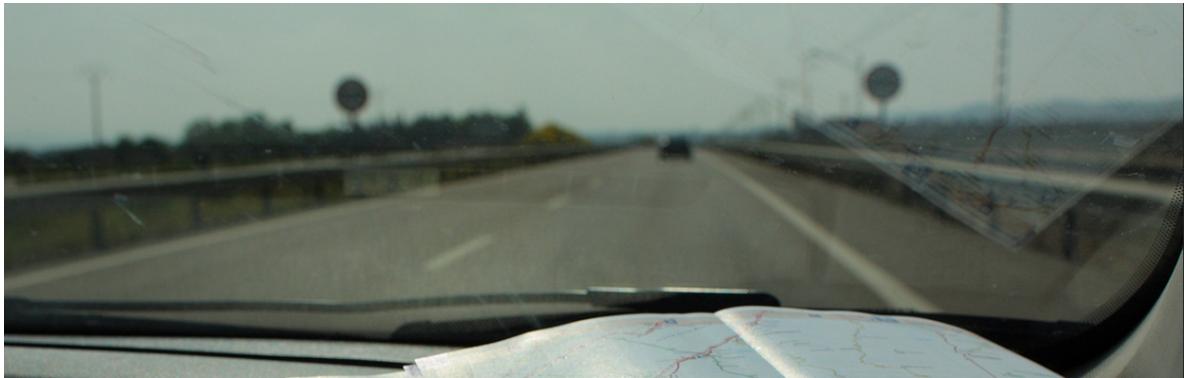
A **geographic information system (GIS)**
is a system designed to

- capture,
- store,
- manipulate,
- analyze,
- manage,
- present

all types of spatial or geographical data.

GIS benefits organizations of all sizes and in almost every industry.

Travel



Search

Not logged in. [Login/Register](#)

PadMapper - Making Apartment Hunting Suck Less.

About • Blue • Eviction • iPhone • Link • Alerts • Jobs • [Add A Pad](#)

Feedback • Email Subscribers • Follow PadMapper on Twitter or Facebook • Sign up

Map Somewhere: Orlando, FL [Go](#)

Filter Listings

Max/Min Rent: \$0 - \$6000+ [Set](#)

of Bedrooms: 0+ - 10+ [Set](#)

Min # Bathrooms: At least 1 BR [Set](#)

Max Price Per Bedroom: At most \$6000+ per BR [Set](#)

Max Age of Listings: At most 7+ days old [Set](#)

Keywords: Words Required in Listing [Set](#) [Clear](#)

Pictures Cats Dogs Owner/No Fee (NYC Only)

Show Full Lease: Show Sub Info:

Sources: Apartments.com ForRent: Others:

Get Email Alerts for this View/Filter Search

Super-Secret Advanced Features

Conserve Time Overlay (Slow) Shows you the rough area you can live in to make your commute less than 30 minutes.

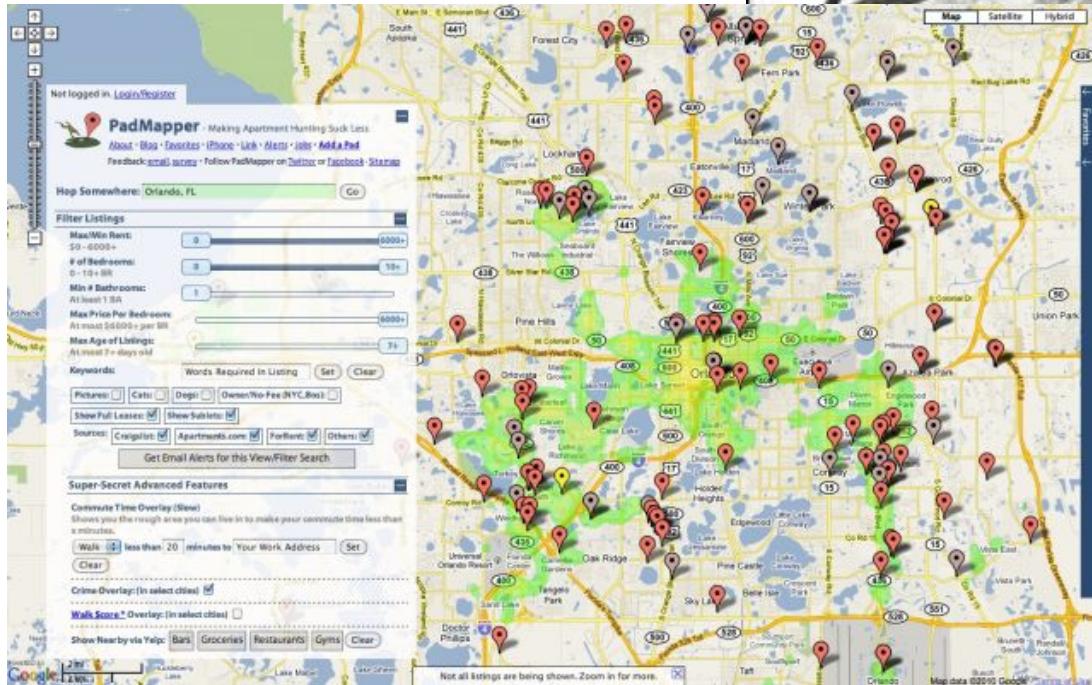
Walk: less than 20+ minutes to Your Work Address [Set](#) [Clear](#)

Crime Overlay: In select cities

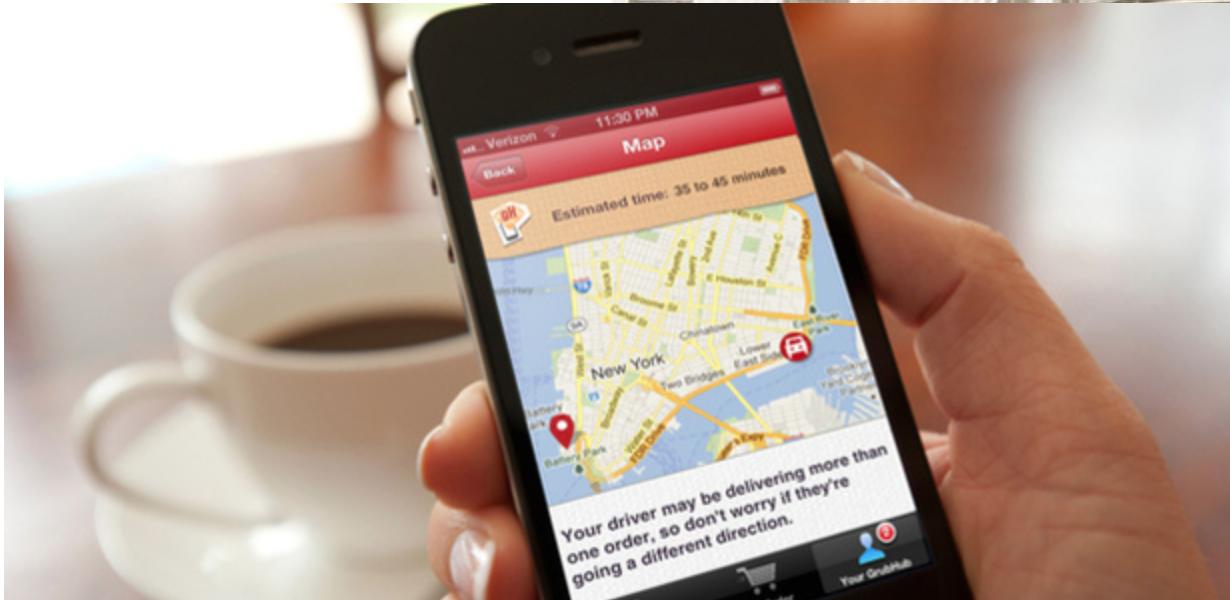
Walk Score™ Overlay: In select cities

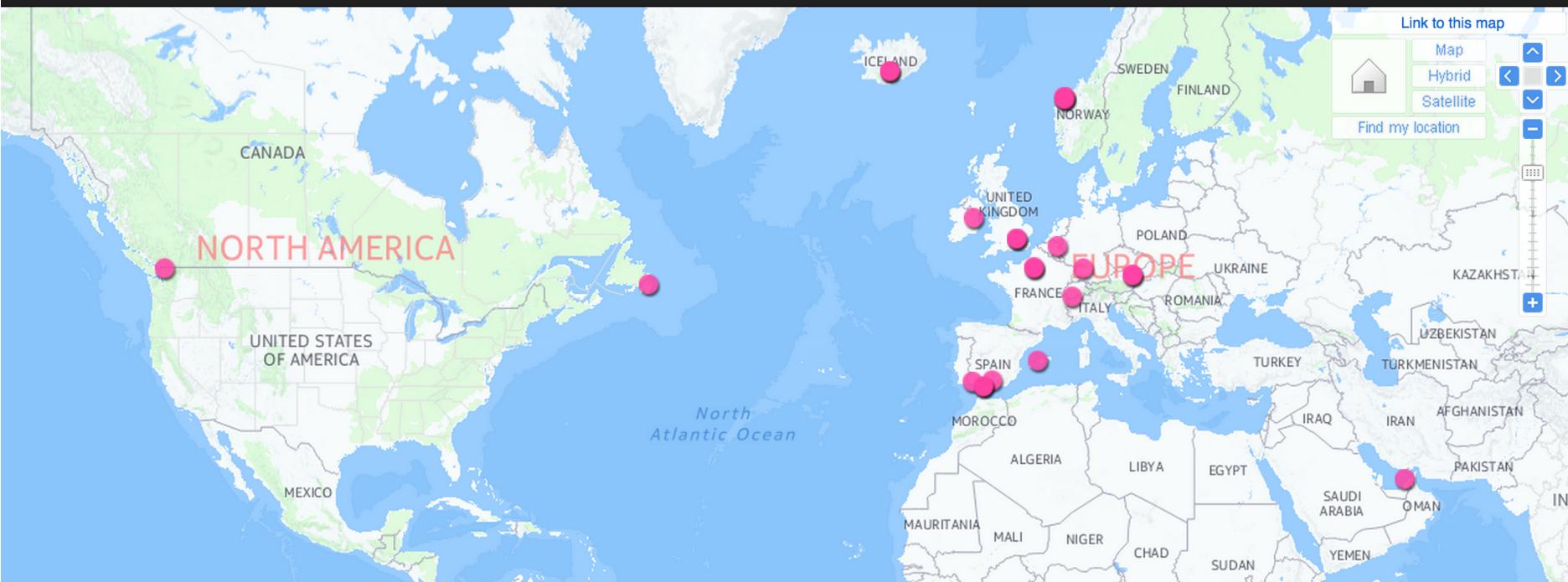
Show Near by via Yelp: Bars • Groceries • Restaurants • Gyms [Clear](#)

Map data ©2009 Google



Logistics & Delivery





2,275,783 geotagged items

Search the map

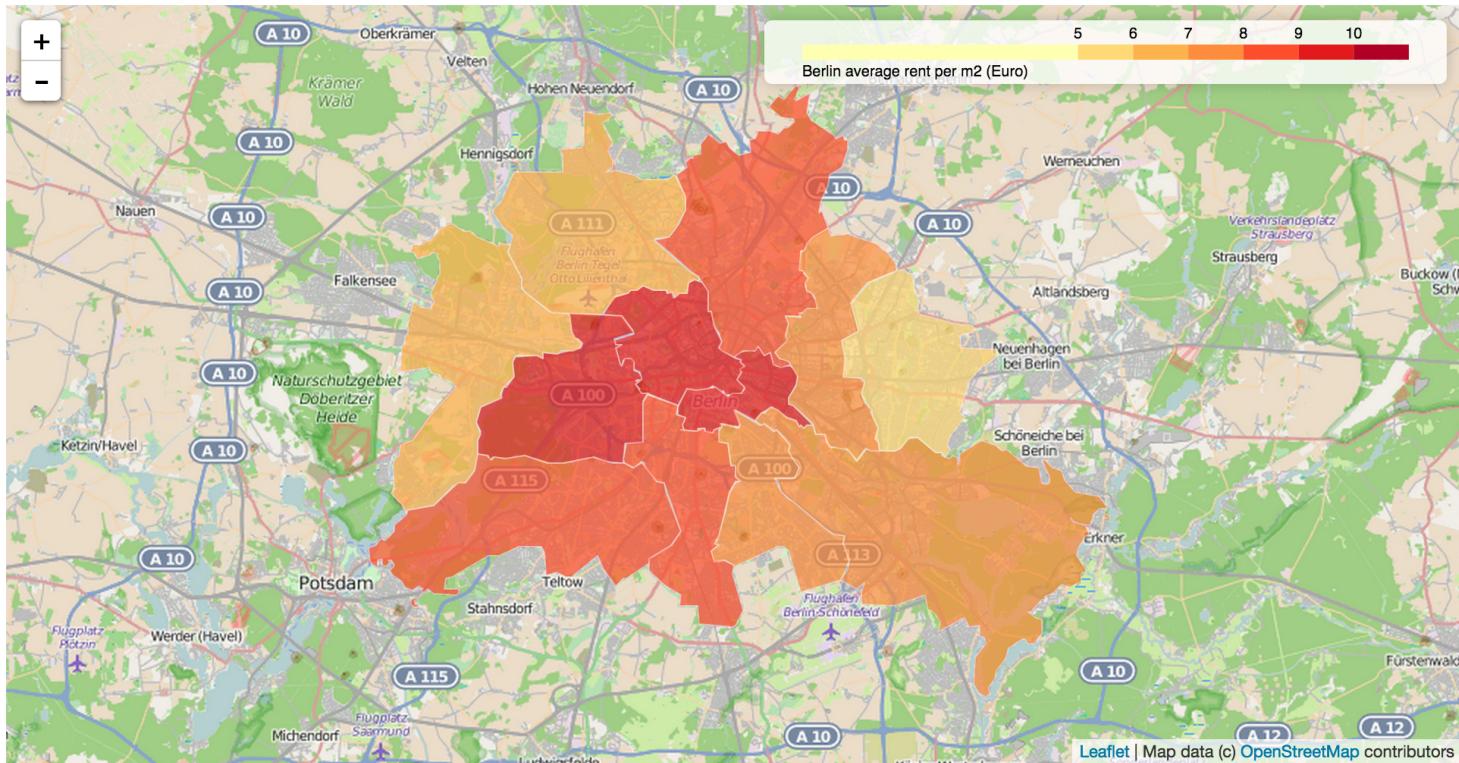
**Location information
is everywhere**

Why map it?

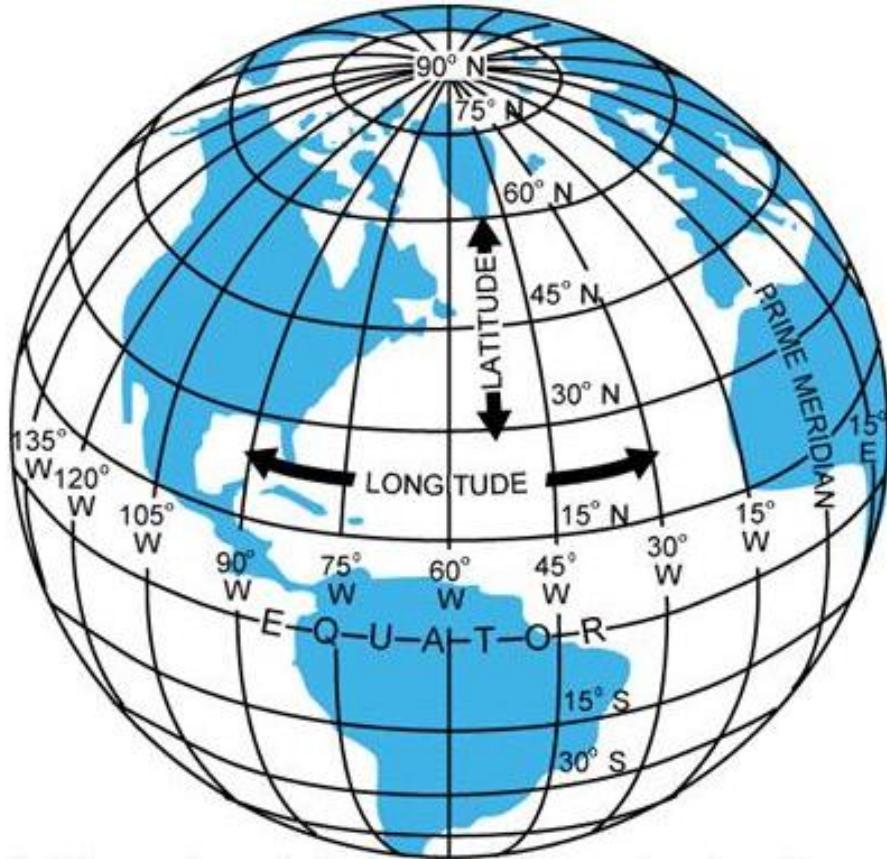
This?

| | zone | avg_rent | population | perc_foreign | pop_thousands | borough | geometry | area_km2 | pop_density |
|---|----------------------------|----------|------------|--------------|---------------|----------------------------|---|------------|-------------|
| 0 | Mitte | 9.36 | 356506 | 30.0 | 357 | Mitte | POLYGON Z ((13.40352847102303 52.5402122900489... | 39.174751 | 9.113013 |
| 1 | Friedrichshain-Kreuzberg | 9.98 | 275691 | 23.7 | 276 | Friedrichshain-Kreuzberg | POLYGON Z ((13.42940239358947 52.5085721176236... | 20.179866 | 13.676999 |
| 2 | Pankow | 8.99 | 384367 | 9.7 | 384 | Pankow | POLYGON Z ((13.38928186179724 52.5682424170027... | 102.357954 | 3.751540 |
| 3 | Charlottenburg-Wilmersdorf | 9.49 | 326354 | 21.9 | 326 | Charlottenburg-Wilmersdorf | POLYGON Z ((13.31278803790569 52.5323108846473... | 64.211650 | 5.076960 |
| 4 | Spandau | 6.58 | 230419 | 14.0 | 230 | Spandau | POLYGON Z ((13.15028090154557 52.5677149012197... | 91.191215 | 2.522173 |
| 5 | Steglitz-Zehlendorf | 8.50 | 299268 | 12.4 | 299 | Steglitz-Zehlendorf | POLYGON Z ((13.34547055014765 52.4152543146776... | 101.790497 | 2.937406 |
| 6 | Tempelhof-Schöneberg | 8.01 | 335767 | 16.7 | 336 | Tempelhof-Schöneberg | POLYGON Z ((13.35597623728278 52.4556212130814... | 52.728447 | 6.372272 |
| 7 | Neukölln | 7.73 | 325716 | 23.5 | 326 | Neukölln | POLYGON Z ((13.4112583396986 | 44.597977 | 7.309749 |

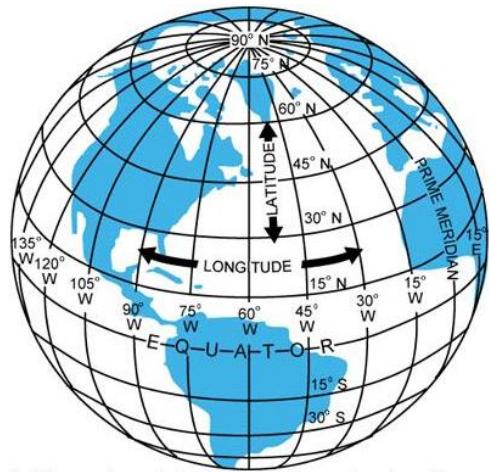
Or this?



Mapping 101: The Earth

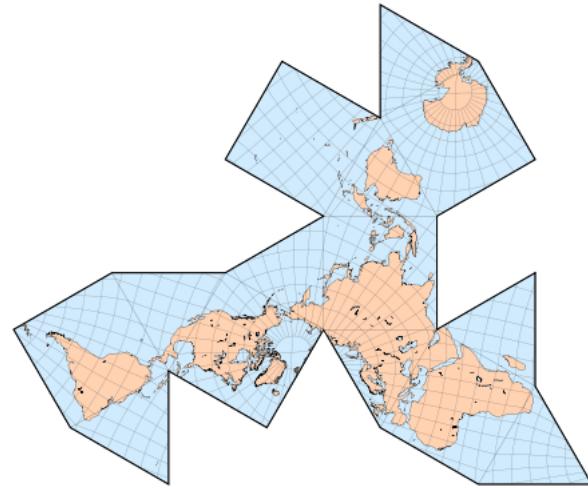


World

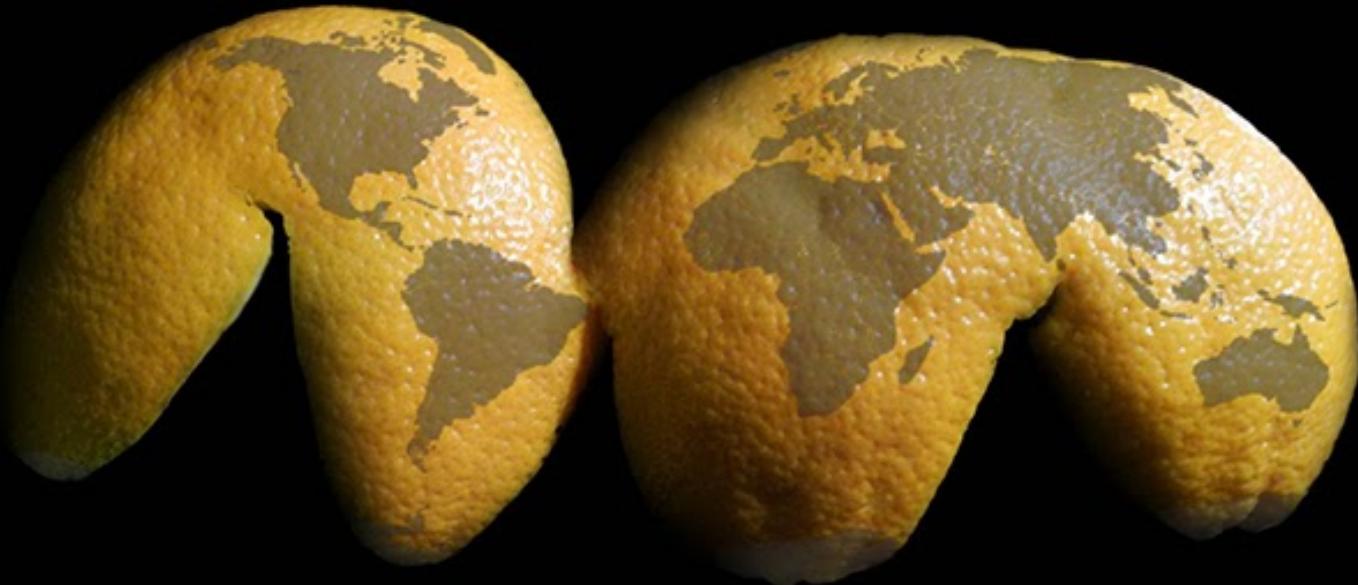


Projection

2D map



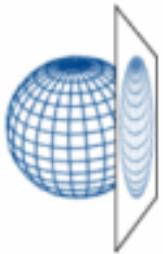
ORANGE PEEL CONTINENTAL GOODE HOMOLOSINE



Nathan P. Belz, Ph.D.
University of Alaska – Fairbanks

© All rights reserved

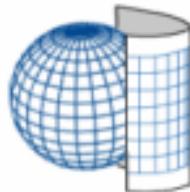
2012.05.15



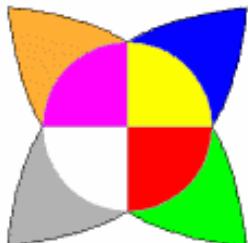
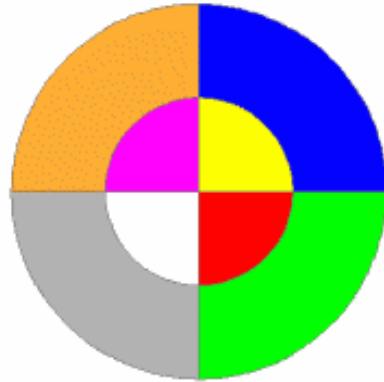
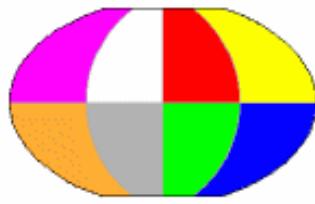
Azimuthal



Conical

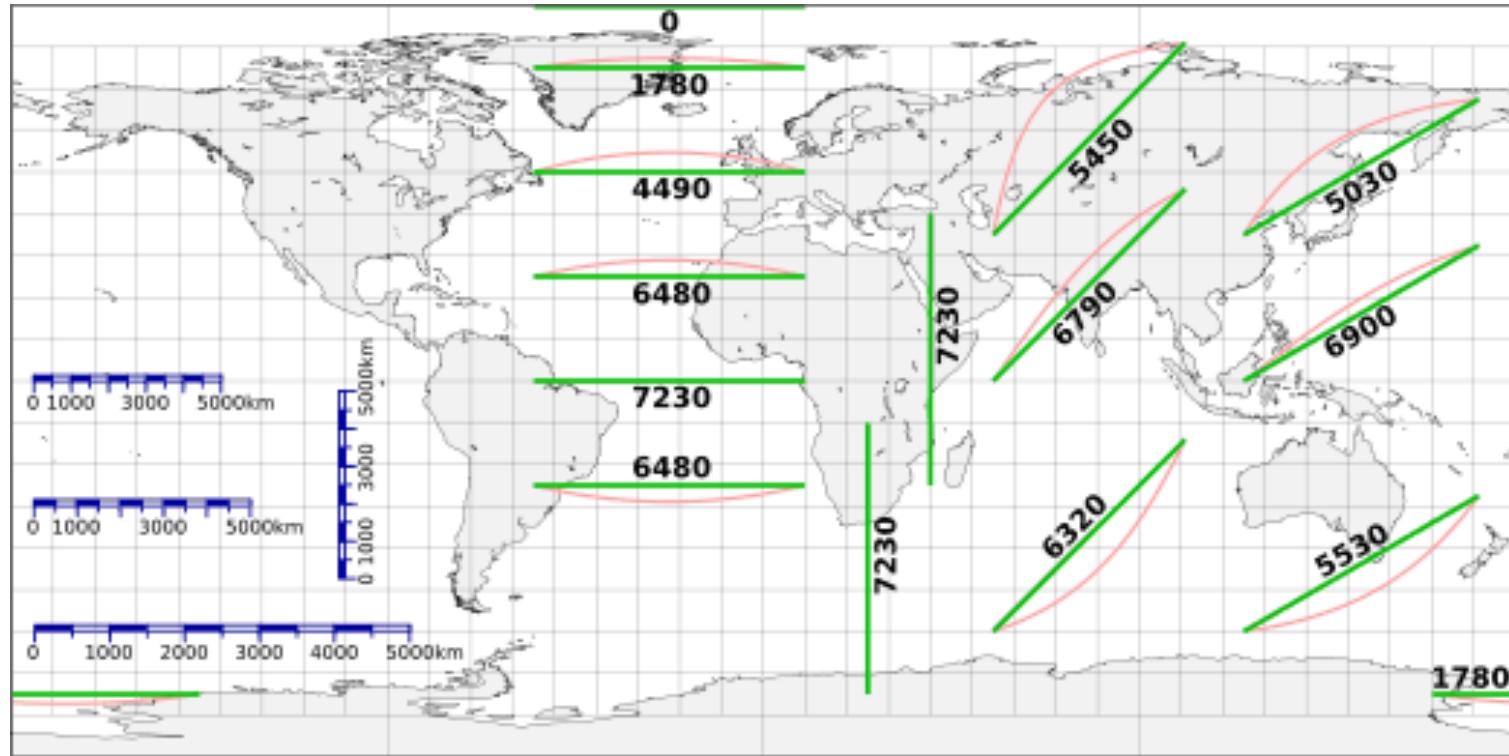


Cylindrical

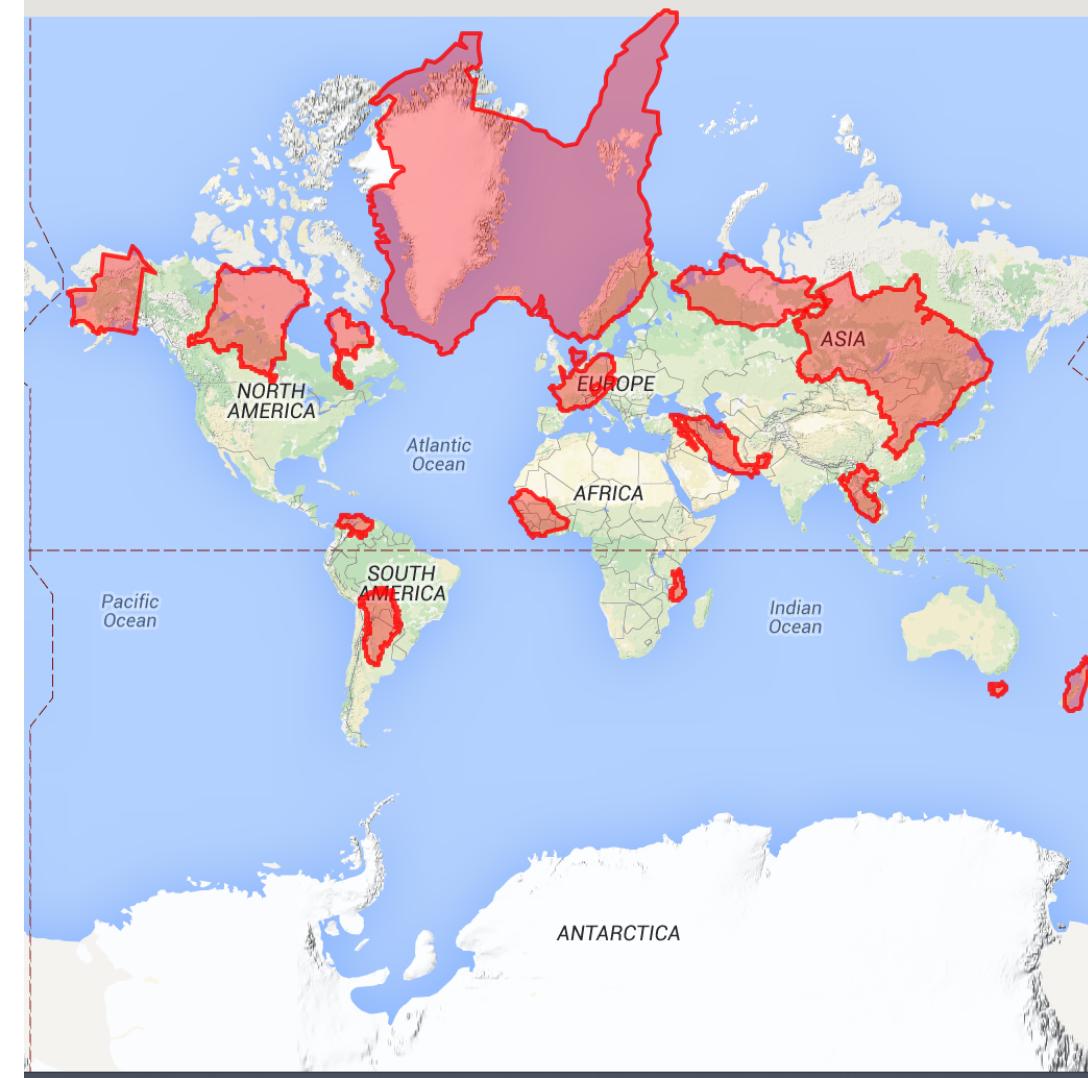


Projections can preserve
area, shape, distance,
angles or other attributes,
but not everything at once

In particular, no flat map can be simultaneously conformal (shape-preserving) and equal-area (area-preserving) in every point.



Distances not preserved



Mercator puzzle

Areas not preserved

WHAT YOUR FAVORITE
MAP PROJECTION
SAYS ABOUT YOU

MERCATOR



YOU'RE NOT REALLY INTO MAPS.

ROBINSON



YOU HAVE A COMFORTABLE PAIR OF RUNNING SHOES THAT YOU WEAR EVERYWHERE. YOU LIKE COFFEE AND ENJOY THE BEATLES. YOU THINK THE ROBINSON IS THE BEST-LOOKING PROJECTION, HANDS DOWN.

WINKEL-TRIPEL



NATIONAL GEOGRAPHIC ADOPTED THE WINKEL-TRIPEL IN 1998, BUT YOU'VE BEEN A WT FAN SINCE LONG BEFORE "Nat Geo" SHOWED UP. YOU'RE WORRIED IT'S GETTING PLAYED OUT, AND ARE THINKING OF SWITCHING TO THE KAVRAYSKY. YOU ONCE LEFT A PARTY IN DISGUST WHEN A GUEST SHOWED UP WEARING SHOES WITH TOES. YOUR FAVORITE MUSICAL GENRE IS "POST".

VAN DER GRINTEN



YOU'RE NOT A COMPLICATED PERSON. YOU LOVE THE MERCATOR PROJECTION; YOU JUST WISH IT WEREN'T SQUARE. THE EARTH'S NOT A SQUARE, IT'S A CIRCLE. YOU LIKE CIRCLES. TODAY IS GONNA BE A GOOD DAY!

Dymaxion



YOU LIKE ISAAC ASIMOV, XML, AND SHOES WITH TOES. YOU THINK THE SEGWAY GOT A BAD RAP. YOU OWN 3D GOGGLES, WHICH YOU USE TO VIEW ROTATING MODELS OF BETTER 3D GOGGLES. YOU TYPE IN DVORAK.

GOODE HOMOLOSINE



THEY SAY MAPPING THE EARTH ON A 2D SURFACE IS LIKE FLATTENING AN ORANGE PEEL, WHICH SEEMS EASY ENOUGH TO YOU. YOU LIKE EASY SOLUTIONS. YOU THINK WE WOULDN'T HAVE SO MANY PROBLEMS IF WE JUST ELECT MORAL PEOPLE TO CONGRESS INSTEAD OF POLITICIANS. YOU THINK AIRLINES SHOULD JUST BUY FOOD FROM THE RESTAURANTS NEAR THE GATES AND SERVE THAT ON BOARD. YOU CHANGE YOUR CAR'S OIL, BUT SECRETLY WONDER IF YOU REALLY NEED TO.

HOB - DYER



YOU WANT TO AVOID CULTURAL IMPERIALISM, BUT YOU'VE HEARD BAD THINGS ABOUT GALL-PETERS. YOU'RE CONFLICT-AVERSE AND BUY ORGANIC. YOU USE A RECENTLY-INVENTED SET OF GENDER-NEUTRAL PRONOUNS AND THINK THAT WHAT THE WORLD NEEDS IS A REVOLUTION IN CONSCIOUSNESS.

A GLOBE!



YES, YOU'RE VERY CLEVER.

PEIRCE QUINCUNCIAL



REALLY? YOU KNOW THE WATERMAN? HAVE YOU SEEN THE 1909 CHILL MAP IT'S BASED - ... YOU HAVE A FRAMED REPRODUCTION AT HOME?! WHOA. LISTEN, FORGET THESE QUESTIONS. ARE YOU DOING ANYTHING TONIGHT?

GALL-PETERS



I HATE YOU.

xkcd 977

Brief GIS overview

Geodata *is* Data

Spatial databases store and manipulate spatial objects



1. **Spatial data types** → shapes such as point, line, and polygon
2. Multi-dimensional **spatial indexing** → efficient processing of spatial operations
3. **Spatial functions** in SQL → querying of spatial properties and relationships.

1. Spatial data types → GEOMETRIES

- **Point**



```
SELECT ST_AsText(geom)
```

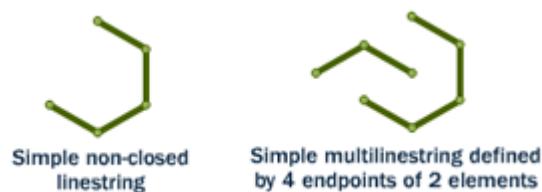
```
FROM geometries
```

```
WHERE name = 'Point';
```

```
-----
```

```
POINT(13.414026 52.502646)
```

- **LineString**

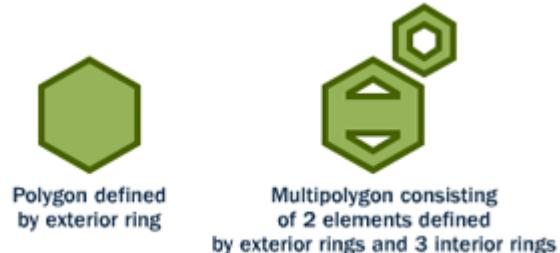


```
WHERE name = 'Linestring';
```

```
-----
```

```
LINESTRING(0 0, 1 1, 2 1, 2 2)
```

- **Polygon**



```
WHERE name = 'Polygon';
```

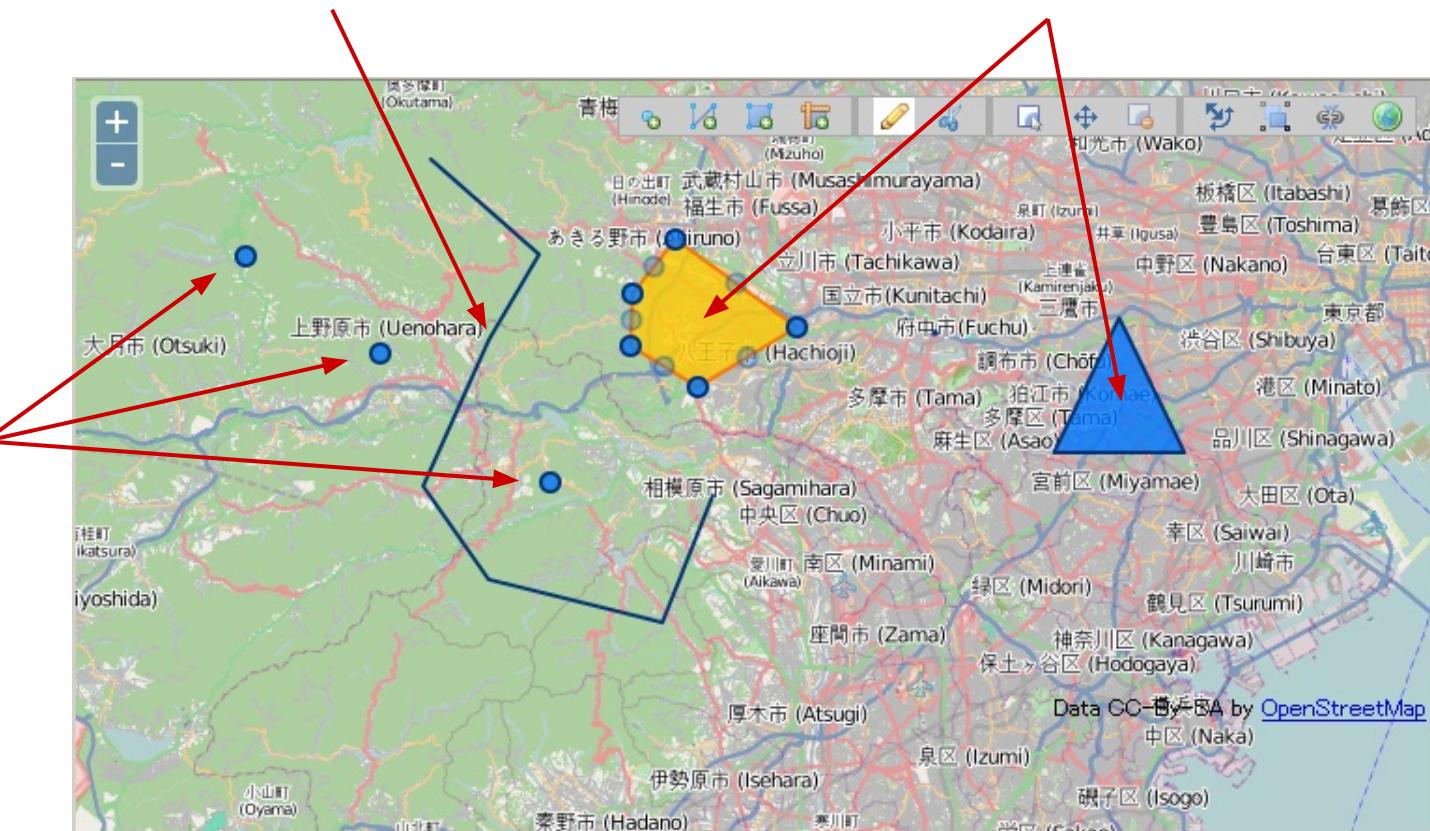
```
-----
```

```
POLYGON((0 0, 1 0, 1 1, 0 1, 0 0))
```

Paths

Areas

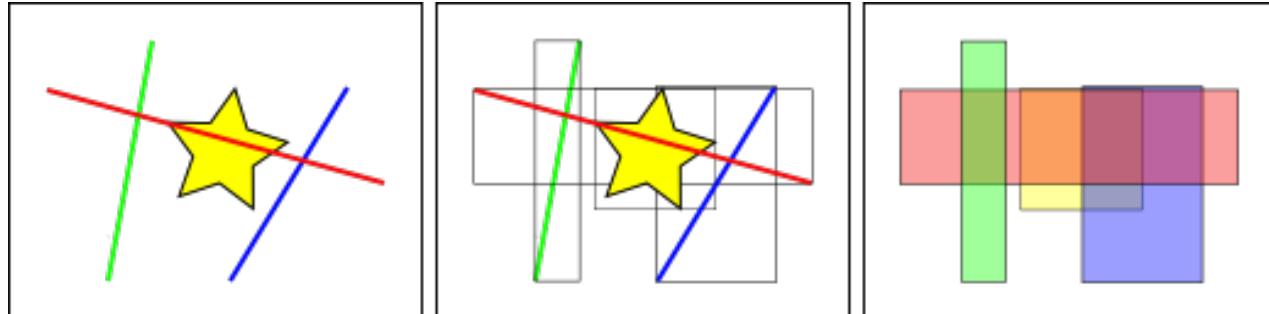
Places



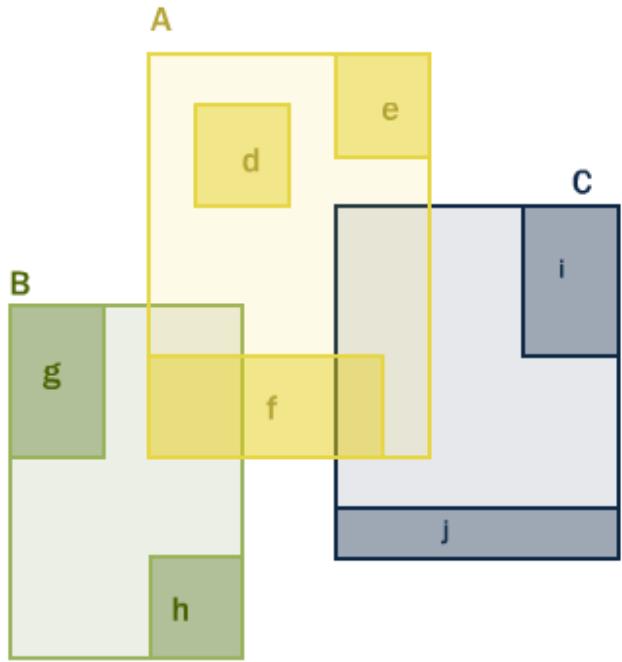
2. Spatial index → EFFICIENT PROCESSING OF OPERATIONS

An ordinary database provides “access methods” (**indexes**)
to allow for fast and random access to subsets of data
→ B-tree* hierarchy [natural sort order puts data into hierarchical tree]

Real spatial databases provide a “**spatial index**” that instead answers the question “which objects are within this particular bounding box?”

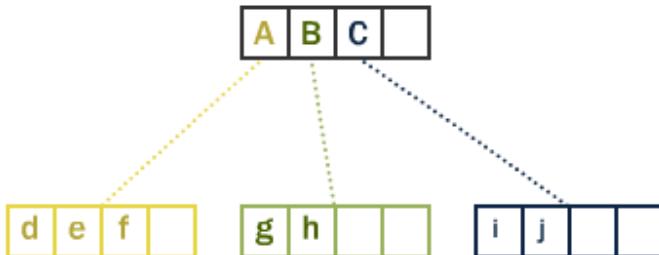


* B-tree ≠ binary tree
Not clear what B stands for
(Boeing, Bayer, balanced...)



R-Tree Hierarchy

- “R” for Rectangle
- Group nearby objects and represent them with their **minimum bounding rectangle** in the next higher level of the tree
- A query that does not intersect the bounding rectangle also cannot intersect any of the contained objects



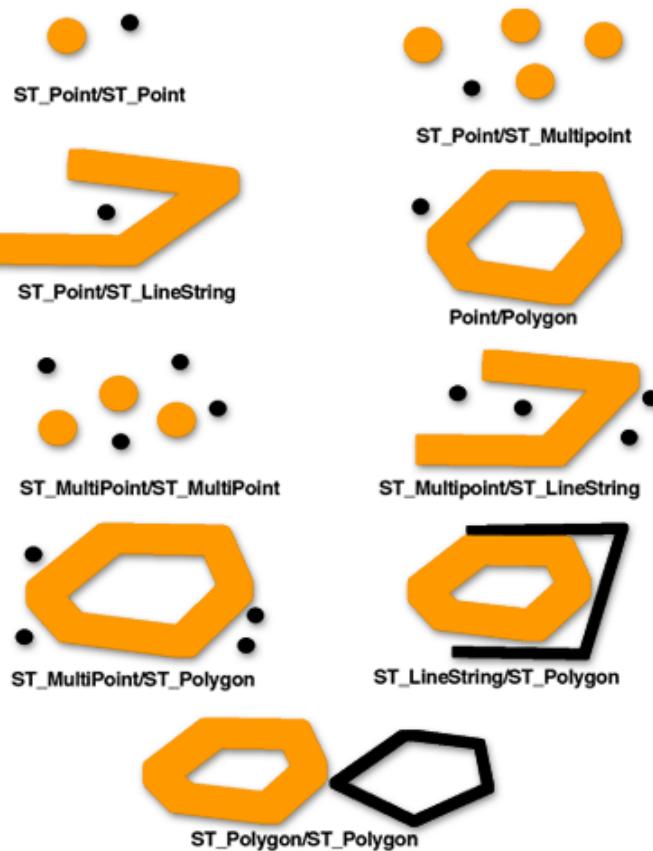
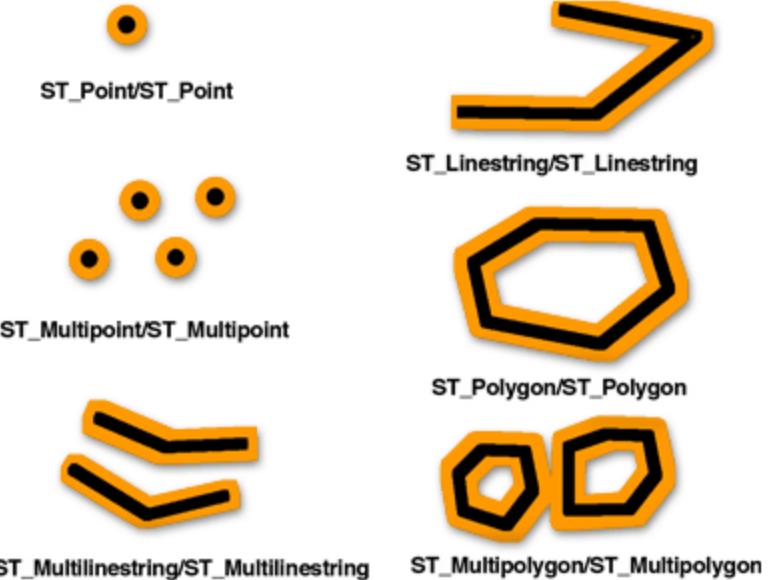
3. **Spatial functions** → SPATIAL PROPERTIES & RELATIONSHIPS

The spatial functions serve as the building block for any spatial project.

The majority of all spatial functions can be grouped into one of the following five categories:

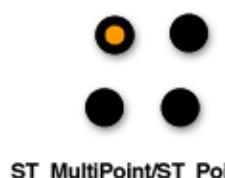
1. **Conversion:** Functions that *convert* between geometries and external data formats.
2. **Management:** Functions that *manage* information about spatial tables and PostGIS administration.
3. **Retrieval:** Functions that *retrieve* properties and measurements of a Geometry.
4. **Comparison:** Functions that *compare* two geometries with respect to their spatial relation.
5. **Generation:** Functions that *generate* new geometries from others.

EQUALS

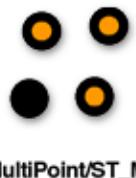


DISJOINT

CONTAINS



ST_MultiPoint/ST_Point



ST_MultiPoint/ST_MultiPoint



ST_LineString/ST_Point



ST_LineString/ST_MultiPoint



ST_Polygon/ST_Point



ST_Polygon/ST_MultiPoint



ST_LineString/ST_LineString



ST_Polygon/ST_LineString



ST_Polygon/ST_Polygon



ST_MultiPoint/ST_LineString



ST_MultiPoint/ST_Polygon



ST_Linestring/ST_Linestring



ST_LineString/ST_Polygon

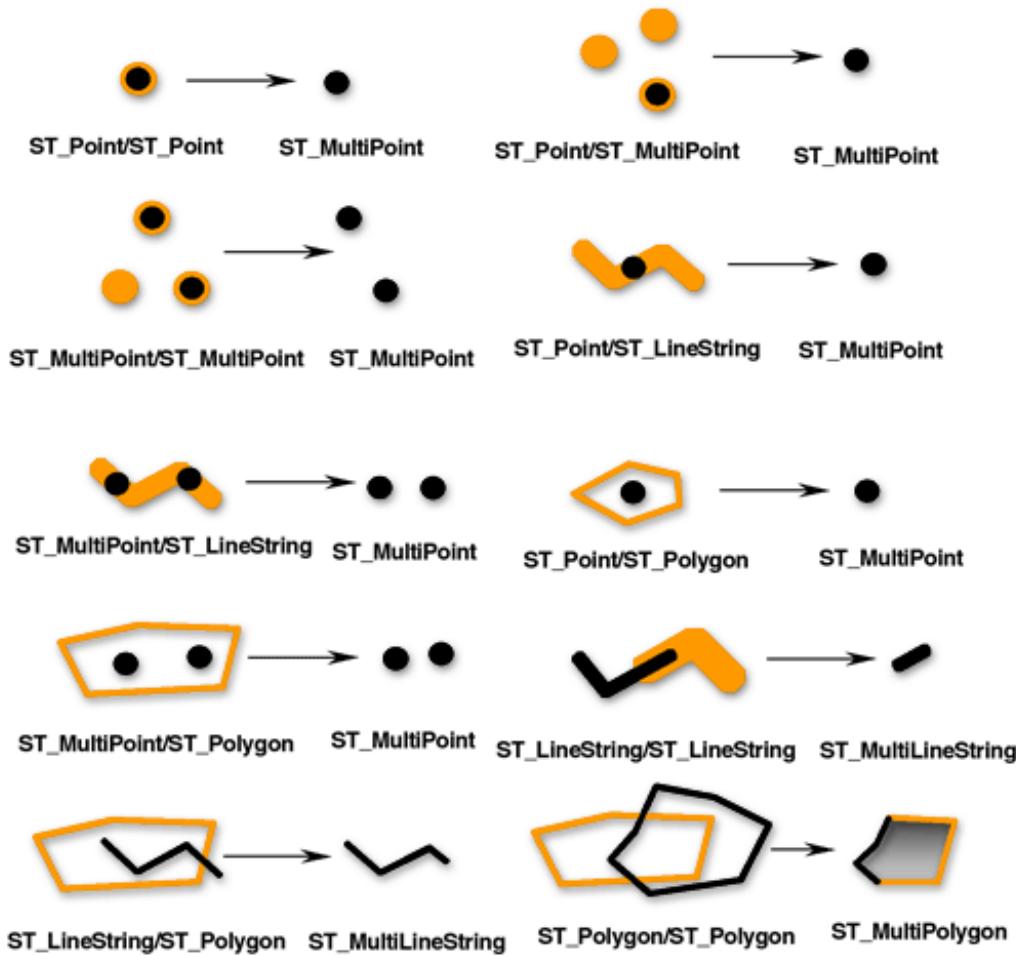
CROSSES

INTERSECTS

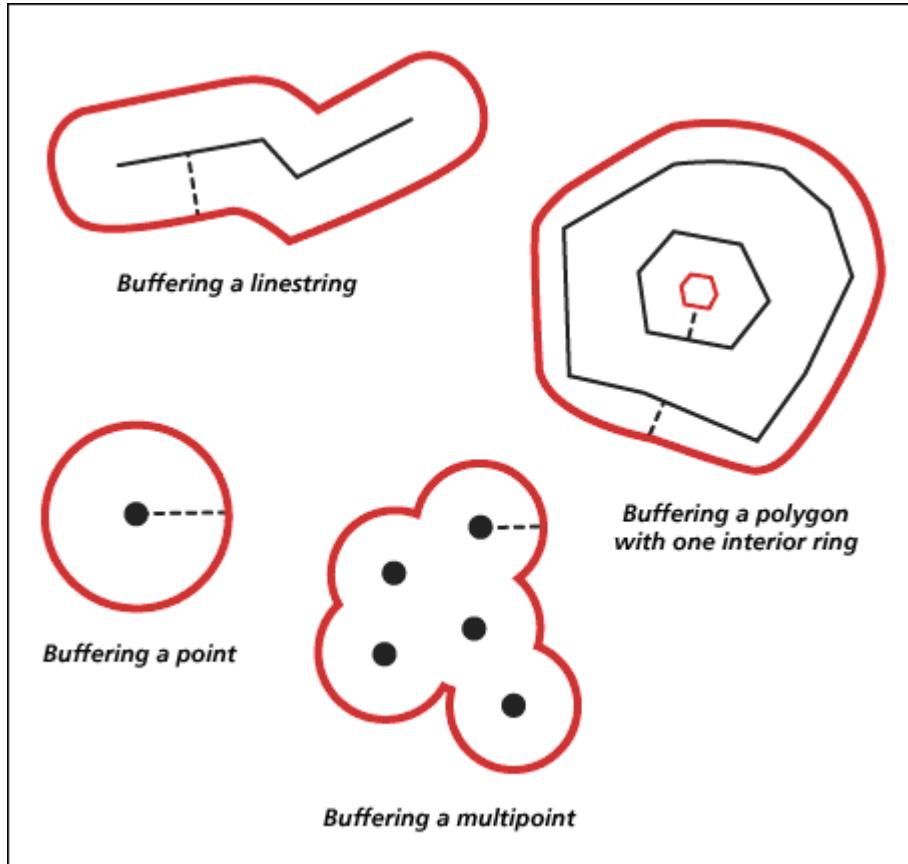
/

INTERSECTION

[if the two shapes have
any space in common,
i.e., if their boundaries
or interiors intersect]



BUFFERING



- Find streets nearby (within X meters of) the subway stop
- Cluster locations
- ...

Hands-on Tutorials

- Basics of 2d geometries and properties
- Easy creation of maps with leaflet/folium

[Next two require creating and loading spatial database:

<https://github.com/lucasantamaria/geodata/blob/master/CreatingLoadingSpatialDB.pdf>]

- Visualization of NYC Census Geospatial data
- Exploring postGIS spatial relationships