

HAS Tools:

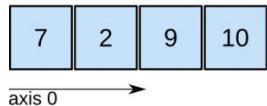
Geospatial data analysis:  
vector data and geopandas

October 7, 2024

# So far we've looked at arrays and data tables as data structures/representations

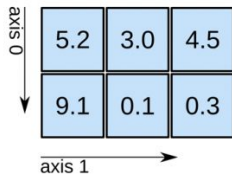
## Array

### 1D array



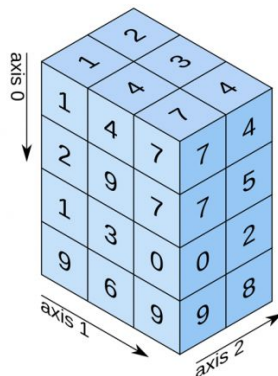
shape: (4,)

### 2D array



shape: (2, 3)

### 3D array



shape: (4, 3, 2)

## DataFrame

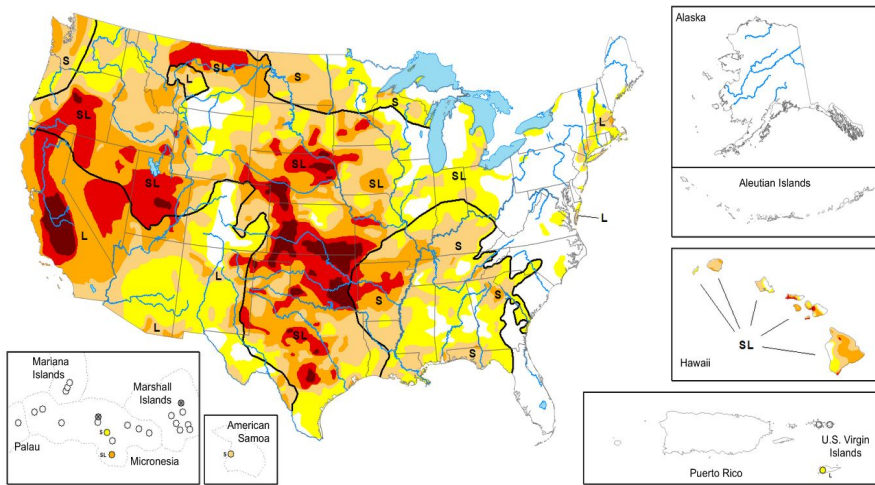
Columns

|   | Name            | Team           | Number | Position | Age  |
|---|-----------------|----------------|--------|----------|------|
| 0 | Avery Bradley   | Boston Celtics | 0.0    | PG       | 25.0 |
| 1 | John Holland    | Boston Celtics | 30.0   | SG       | 27.0 |
| 2 | Jonas Jerebko   | Boston Celtics | 8.0    | PF       | 29.0 |
| 3 | Jordan Mickey   | Boston Celtics | NaN    | PF       | 21.0 |
| 4 | Terry Rozier    | Boston Celtics | 12.0   | PG       | 22.0 |
| 5 | Jared Sullinger | Boston Celtics | 7.0    | C        | NaN  |
| 6 | Evan Turner     | Boston Celtics | 11.0   | SG       | 27.0 |

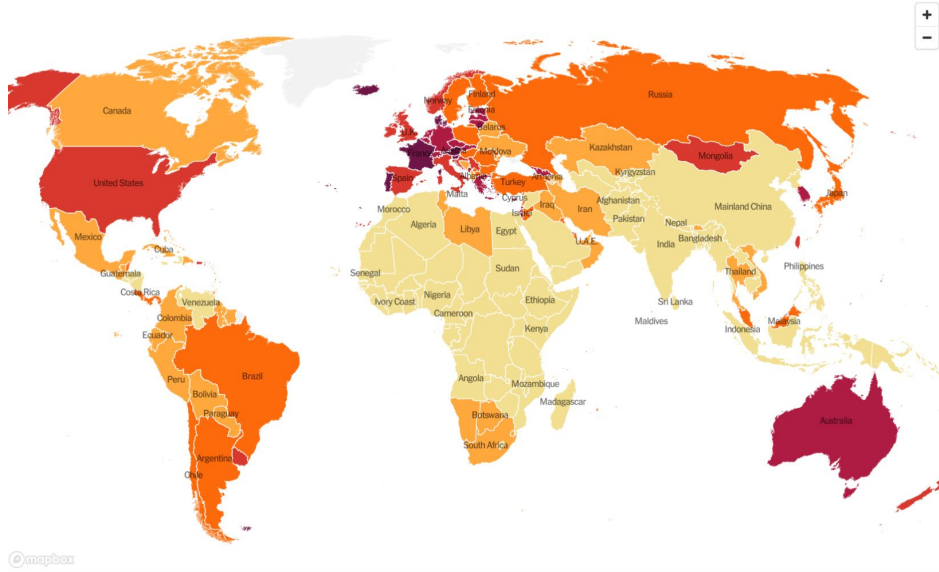
Rows

Data

# But map making and geospatial analysis is important in Earth & environmental science



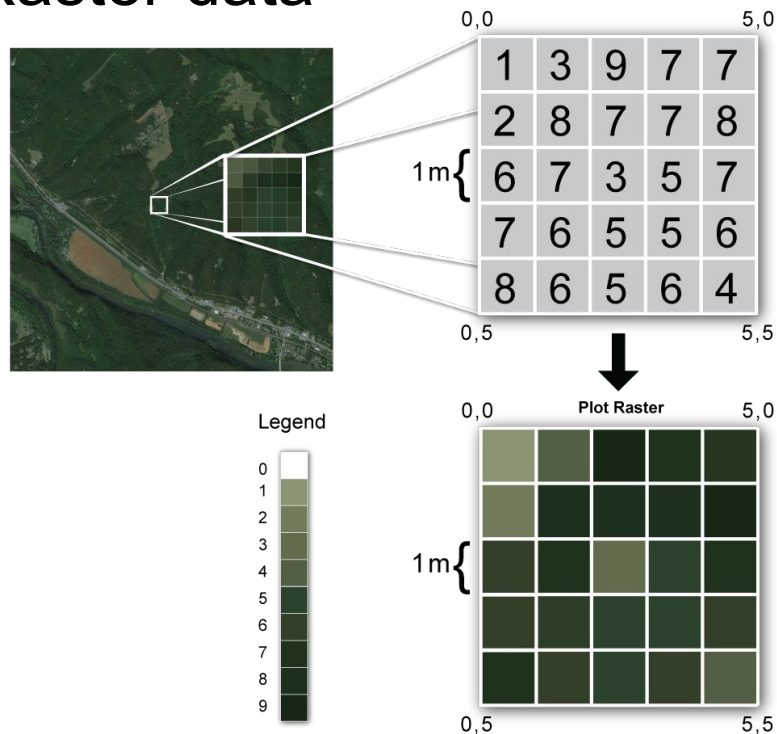
<https://droughtmonitor.unl.edu/>



<https://www.nytimes.com/interactive/2021/world/covid-cases.html>

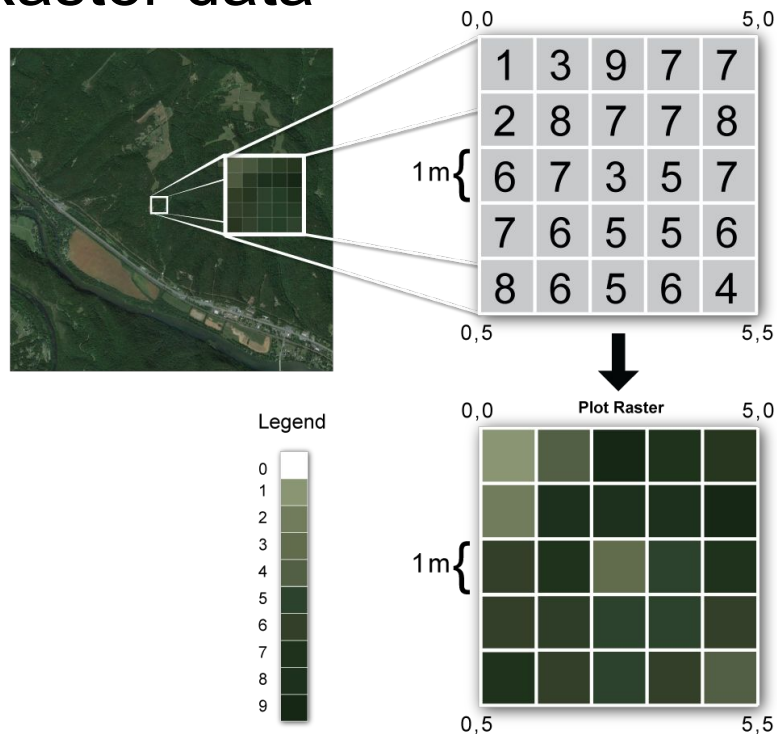
# Geographic Information Systems (GIS) have 2 main data representations

## Raster data



# Geographic Information Systems (GIS) have 2 main data representations

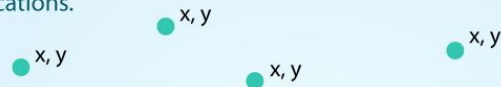
## Raster data



## Vector data

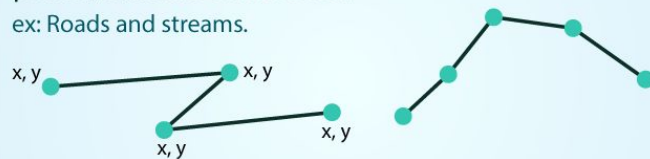
**POINTS:** Individual  $x, y$  locations.

ex: Center point of plot locations, tower locations, sampling locations.



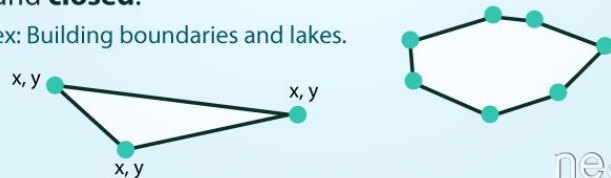
**LINES:** Composed of many (at least 2) vertices, or points, that are connected.

ex: Roads and streams.



**POLYGONS:** 3 or more vertices that are connected and **closed**.

ex: Building boundaries and lakes.

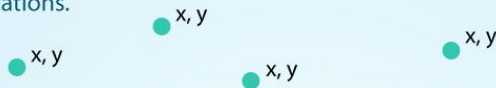


neon

# Today we begin to investigate vector data

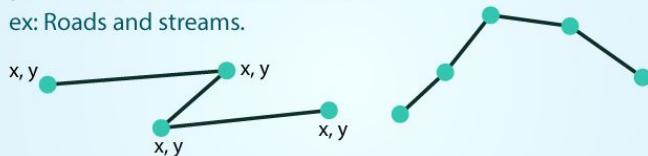
**POINTS:** Individual  $x, y$  locations.

ex: Center point of plot locations, tower locations, sampling locations.



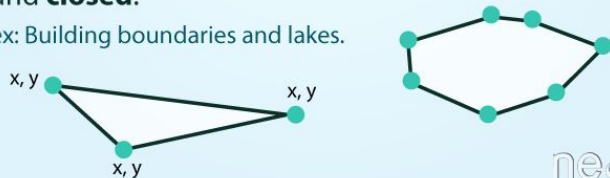
**LINES:** Composed of many (at least 2) vertices, or points, that are connected.

ex: Roads and streams.

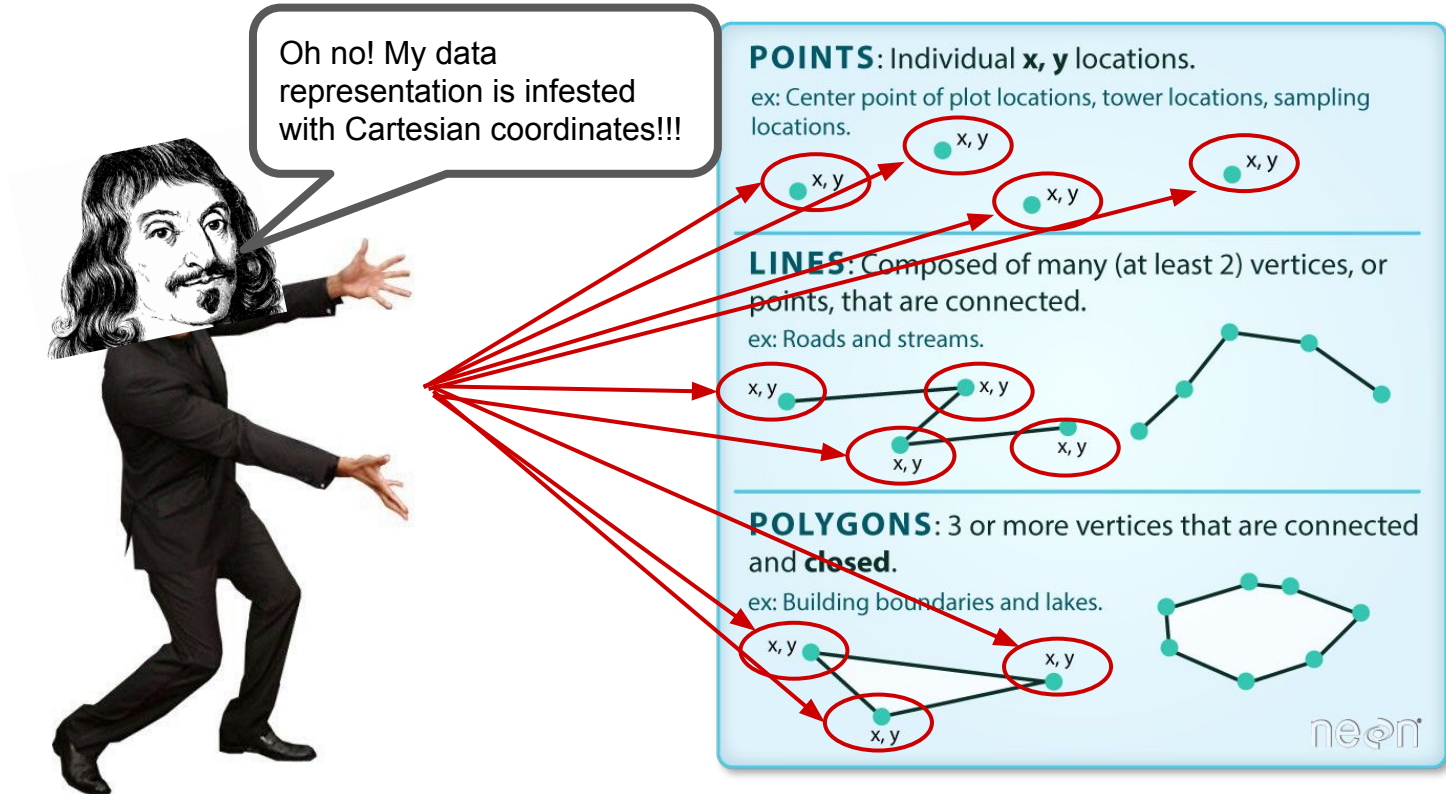


**POLYGONS:** 3 or more vertices that are connected and **closed**.

ex: Building boundaries and lakes.



# One problem though... Earth isn't flat



Oh no! My data representation is infested with Cartesian coordinates!!!

**POINTS:** Individual  $x, y$  locations.  
ex: Center point of plot locations, tower locations, sampling locations.

**LINES:** Composed of many (at least 2) vertices, or points, that are connected.  
ex: Roads and streams.

**POLYGONS:** 3 or more vertices that are connected and **closed**.  
ex: Building boundaries and lakes.

neon





OR IS IT???





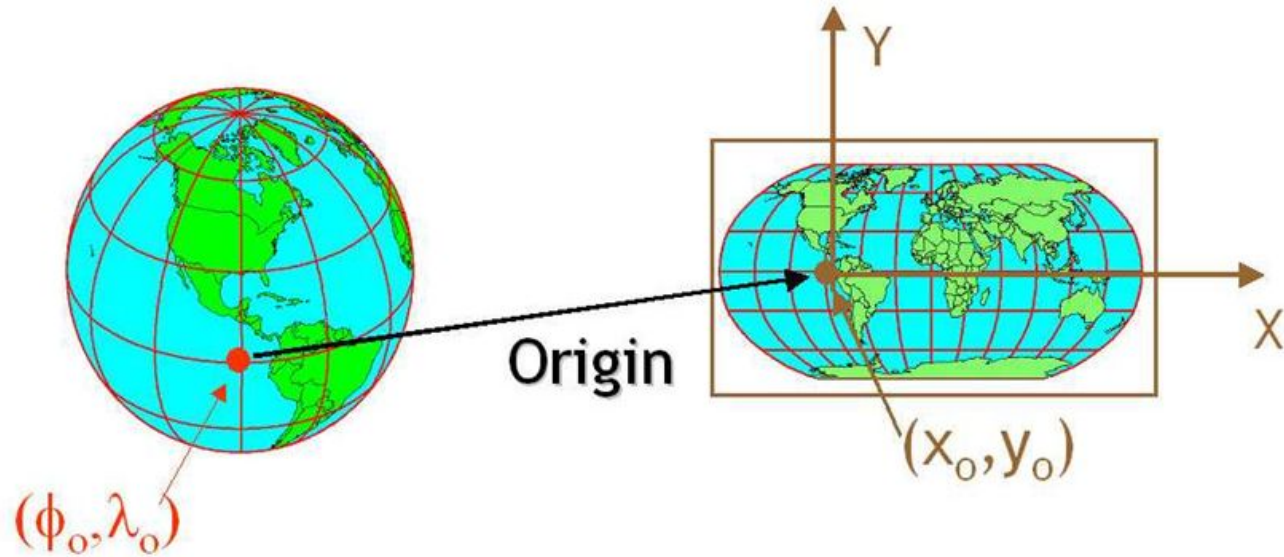
OR IS IT???

**NO**

Fun fact! It is mathematically impossible to map a sphere onto a plane with no distortion.



To get around this we can “project” the sphere onto the plane while trying to preserve certain properties...



# Many ways to do this, and of course, there is an XKCD for it.

WHAT YOUR FAVORITE  
MAP PROJECTION  
SAYS ABOUT YOU

MERCATOR



YOU'RE NOT REALLY INTO MAPS.

ROBINSON



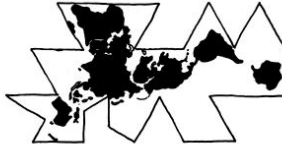
YOU HAVE A COMFORTABLE PAIR OF RUNNING SHOES THAT YOU WEAR EVERYWHERE. YOU LIKE COFFEE AND ENJOY THE BEATLES. YOU THINK THE ROBINSON IS THE BEST-LOOKING PROJECTION, HANDS DOWN.

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. TORY'S GONNA BE A GOOD DAY!

DYMAXION



YOU LIKE ISAAC ASIMOV, XML, AND SHOES WITH TIES. 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.

WINKEL-TRIPEL



NATIONAL GEOGRAPHIC ADOPTED THE WINKEL-TRIPEL IN 1998, BUT YOU'VE BEEN A WWT 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 TIES. YOUR FAVORITE MUSICAL GENRE IS "POST-".

HOBBO-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.

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'D JUST ELECT *NORMAL* PEOPLE TO CONGRESS INSTEAD OF POLITICIANS. YOU THINK AIRLINES SHOULD JUST BUY ROOF 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.

PLATE CARRÉE  
(EQUIRECTANGULAR)



YOU THINK THIS ONE IS FINE. YOU LIKE HOW X AND Y MAP TO LATITUDE AND LONGITUDE. THE OTHER PROJECTIONS OVERCOMPLICATE THINGS. YOU WANT ME TO STOP ASKING ABOUT MAPS SO YOU CAN ENJOY DINNER.

A GLOBE!



YES, YOU'RE VERY CLEVER.

PEIRCE QUINCUNCIAL



YOU THINK THAT WHEN WE LOOK AT A MAP, WHAT WE REALLY SEE IS OURSELVES. AFTER YOU FIRST SAW *INCEPTION*, YOU SAT SILENT IN THE THEATER FOR SIX HOURS. IT BREAKS YOU OUT TO REALIZE THAT EVERYONE AROUND YOU HAS A SKELETON INSIDE THEM. YOU *HAVE* REALLY LOOKED AT YOUR HANDS.

WATERMAN BUTTERFLY



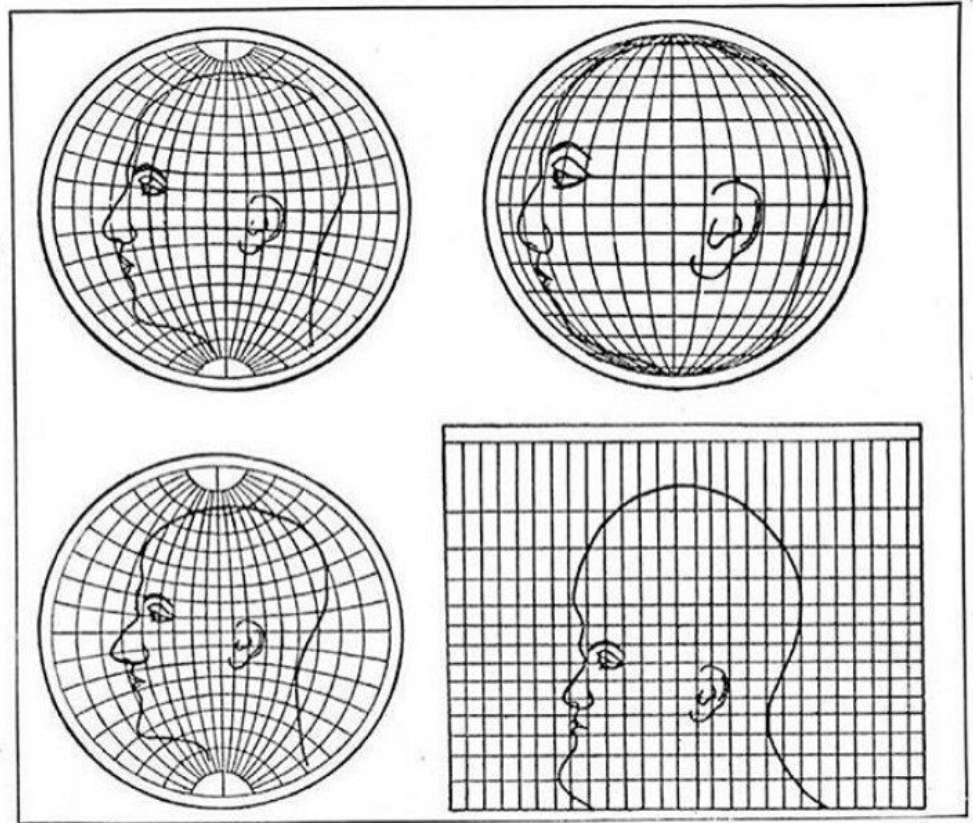
REALLY? YOU KNOW THE WATERMAN? HAVE YOU SEEN THE 1909 CAHILL MAP IT'S BASED - ... YOU HAVE A FRAMED REPRODUCTION AT HOME?! WHOA ... LISTEN, FORGET THESE QUESTIONS. ARE YOU DOING ANYTHING TONIGHT?

GALL-PETERS



I HATE YOU.

There's also this gem,  
although I couldn't find  
the original citation,  
unfortunately.



*Upper left: Globular. Upper right: Orthographic. Lower left: Stereographic.  
Lower right: Mercator*

**What four commonly used projections do, as shown on a human head**



Kudos to Aldo for also  
providing this effective  
illustration of...  
something?





# Some things you should be aware of

## GDAL / OGR

### Geospatial Data Abstraction Library.

- The swiss army knife for geospatial.
- Read and write Raster (GDAL) and Vector (OGR) datasets
- More than 200 (mainly) geospatial formats and protocols.



<https://archive.fosdem.org/2018/schedule/event/geopandas/attachments/slides/2487/export/events/attachments/geopandas/slides/2487/slides.pdf>

# Some things you should be aware of

**GEOS**

**GEOS**

Geometry  
Engine  
Open  
Source

## Geometry Engine Open Source

- C/C++ port of a subset of Java Topology Suite (JTS)
- Most widely used geospatial C++ geometry library
- Implements geometry objects (simple features), spatial predicate functions and spatial operations

Used under the hood by many applications (QGIS, PostGIS, MapServer, GRASS, GeoDjango, ...)

[geos.osgeo.org](https://geos.osgeo.org)

<https://archive.fosdem.org/2018/schedule/event/geopandas/attachments/slides/2487/export/events/attachments/geopandas/slides/2487/slides.pdf>

# Geospatial packages for python

As always, this space is large and we're only covering a snippet

- Raw bindings for GDAL/OGR: <https://pypi.org/project/osgeo/>
- Better bindings for GDAL/OGR:
  - Rasterio for GDAL: <https://rasterio.readthedocs.io/en/latest/>
  - Fiona for OGR: <https://fiona.readthedocs.io/en/latest/>
- Higher level support for vector/geometric analysis:  
<https://shapely.readthedocs.io/en/stable/manual.html>
- High level support for merging data and geometries:  
<https://geopandas.org/en/stable/>
- The best way to work with general raster  
(and other labeled multi-dimensional arrays): <https://docs.xarray.dev/en/stable/>



Next time we will explore  
how this all works with  
GeoPandas, but for now,  
let's play around with  
projections



