An introduction to geospatial data & processing

Dr Robin Wilson

www.rtwilson.com

What is geospatial data?

Schools in Oxfordshire

ID	Name	Type	Status	Phase	Gender	Pupils
123290	Headington School	Other Independent School	Open	Secondary	Girls	1046
122967	Bartlemas Nursery School	LA Nursery School	Open	Nursery	Mixed	63
122968	Headington Quarry Foundation Stage School	LA Nursery School	Open	Nursery	Mixed	81
122969	Grandpont Nursery School	LA Nursery School	Open	Nursery	Mixed	75
123129	Dorchester St Birinus Church of England School	Voluntary Controlled School	Open	Primary	Mixed	78
123130	Great Milton Church of England Primary School	Voluntary Controlled School	Open	Primary	Mixed	161
123131	Marsh Baldon Church of England Controlled School	Voluntary Controlled School	Open	Primary	Mixed	70
123132	Culham Parochial Church of England School	Voluntary Controlled School	Open	Primary	Mixed	51
123133	Crowmarsh Gifford Church of England School	Voluntary Controlled School	Open	Primary	Mixed	204
123134	Harpsden Parochial School	Voluntary Controlled School	Closed	Primary	Mixed	0

Data

Schools in Oxfordshire

ID	Name	Type	Status	Phase	Gender	Pupils	X	Y
123290	Headington School	Other Independent School	Open	Secondary	Girls	1046	453825	206828
122967	Bartlemas Nursery School	LA Nursery School	Open	Nursery	Mixed	63	453107	205197
122968	Headington Quarry Foundation Stage School	LA Nursery School	Open	Nursery	Mixed	81	455495	206950
122969	Grandpont Nursery School	LA Nursery School	Open	Nursery	Mixed	75	451204	205180
123129	Dorchester St Birinus Church of England School	Voluntary Controlled School	Open	Primary	Mixed	78	457895	194371
123130	Great Milton Church of England Primary School	Voluntary Controlled School	Open	Primary	Mixed	161	462976	202879
123131	Marsh Baldon Church of England Controlled School	Voluntary Controlled School	Open	Primary	Mixed	70	456387	199612
123132	Culham Parochial Church of England School	Voluntary Controlled School	Open	Primary	Mixed	51	450766	195061
123133	Crowmarsh Gifford Church of England School	Voluntary Controlled School	Open	Primary	Mixed	204	461707	189029
123134	Harpsden Parochial School	Voluntary Controlled School	Closed	Primary	Mixed	0	475602	180862

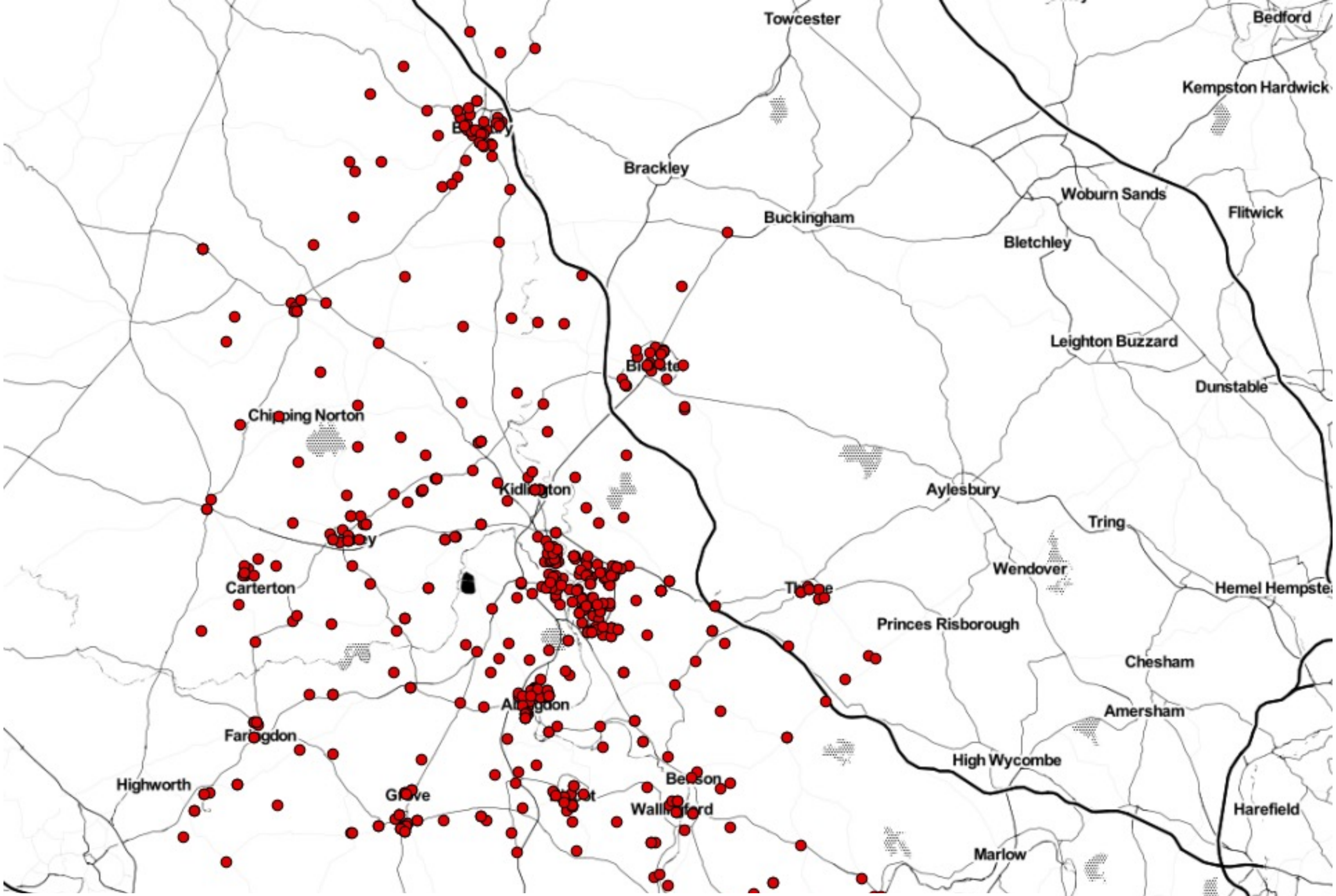
Geographic Data (‘Geodata’)

Schools in Oxfordshire

Everything happens somewhere

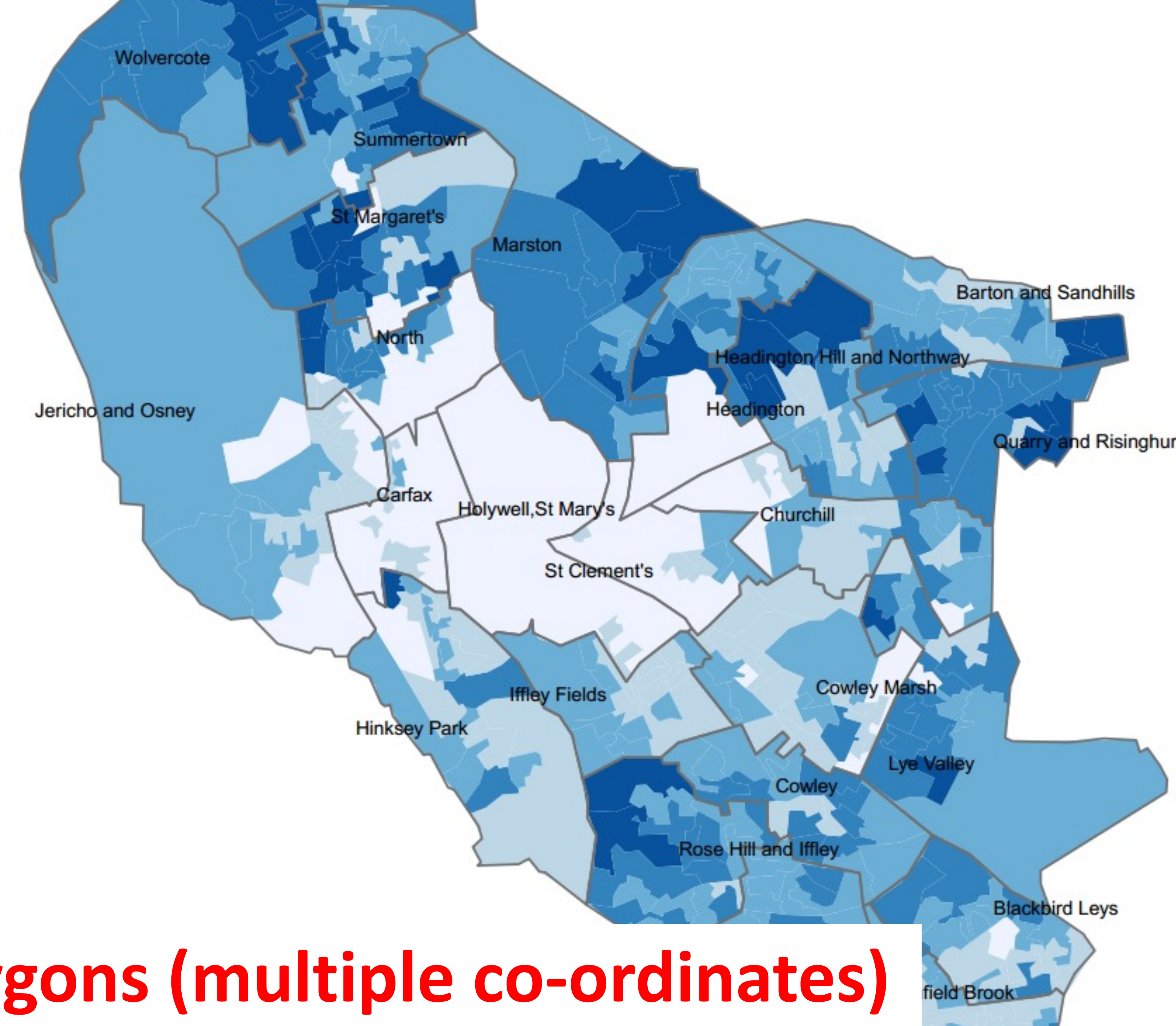
ID	Name	T[illegible]e	S[illegible]u[illegible]	[illegible]has[illegible]	Ge[illegible]er	[illegible]	X	Y
123290	Headington S[illegible]ol	O[illegible]r In[illegible]pende[illegible] School	[illegible]pen	Sec[illegible]ary	Gi[illegible]	1[illegible]	453825	206828
122967	Bartlemas Nu[illegible]	[illegible] Nurse[illegible] School	Open	Nursery	Mi[illegible]ed	[illegible]	453107	205197
122968	Headington Quarry Foundation Stage School	LA Nursery School	Open	Nursery	Mixed	81	455495	206950
122969	Grandpont Nursery School	LA Nursery School	Open	Nursery	Mixed	75	451204	205180
123129	Dorchester St Birinus Ch[illegible]h of England School	Voluntary Controlled School	Open	Primary	Mixed	78	457895	194371
123130	Great Milton Church of E[illegible]ma[illegible] School	V[illegible]ntar[illegible]ed S[illegible]ol	[illegible]en	P[illegible]	[illegible]ixed	161	462976	202879
123131	Marsh Baldon Church of [illegible]glan[illegible] Controlled School	V[illegible]ntar[illegible]ed School	[illegible]en	Primary	Mixed	70	456387	199612
123132	Culham Parochial Church of England School	Voluntary Controlled School	Open	Primary	Mixed	51	450766	195061
123133	Crowmarsh Gifford Church of England School	Voluntary Controlled School	Open	Primary	Mixed	204	461707	189029
123134	Harps[illegible] Paroch[illegible] Schoo[illegible]	V[illegible]tary C[illegible]lled Sch[illegible]	[illegible]se[illegible]	Pr[illegible]ary	[illegible]xe[illegible]	0	47[illegible]02	180862

Geographic Data ('Geodata')

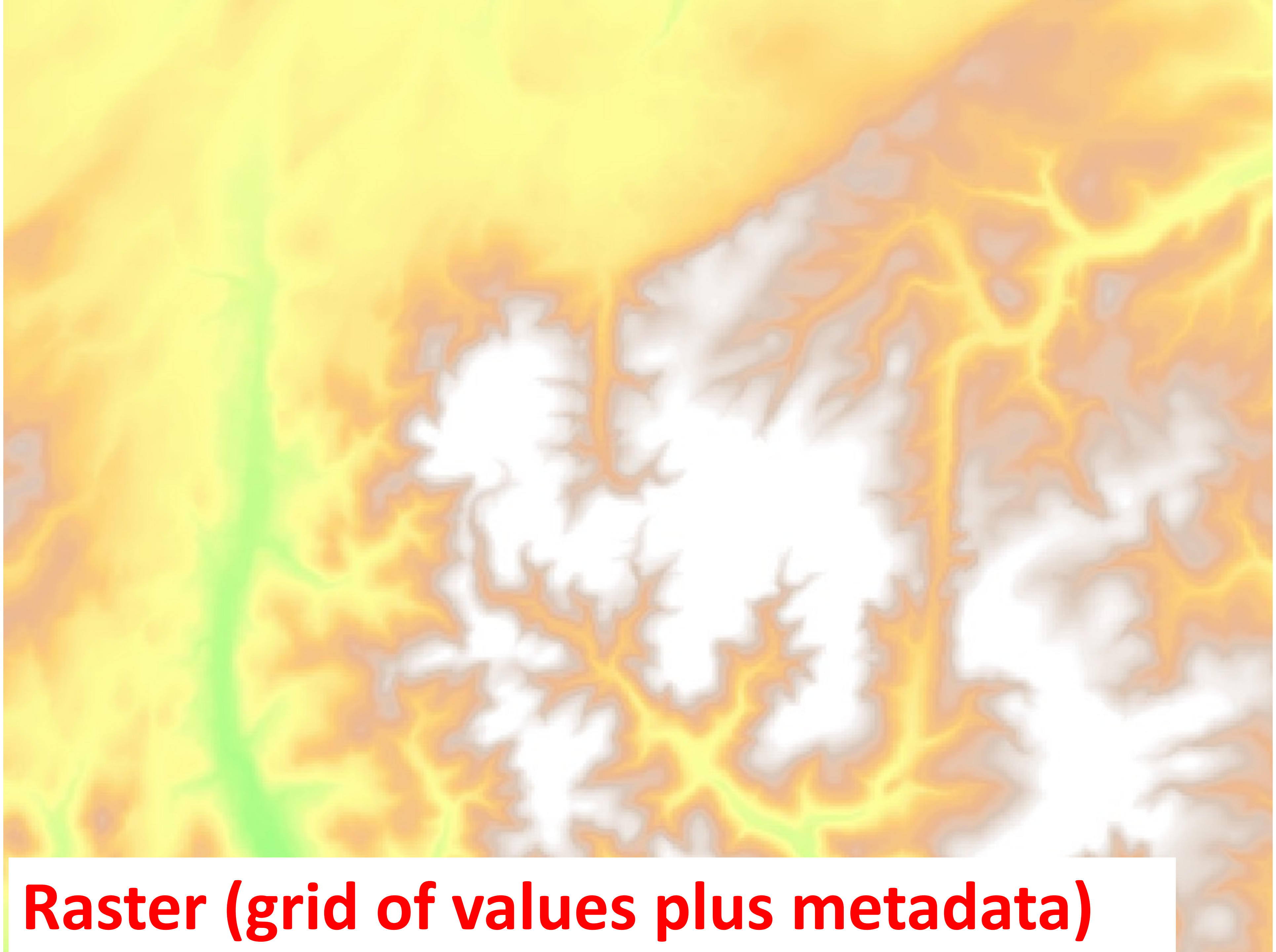


Points (one co-ordinate: X and Y)

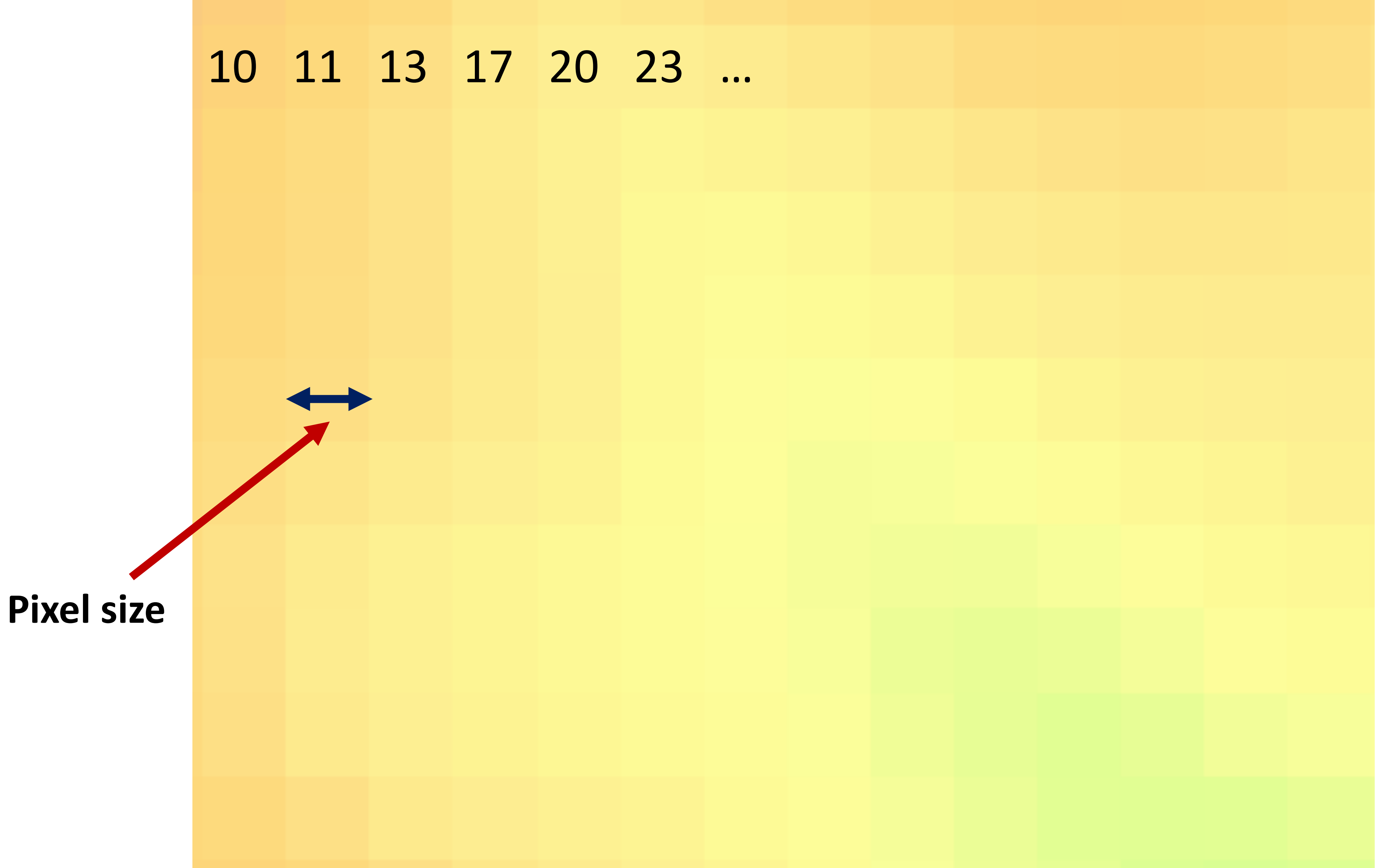
Lines (multiple co-ordinates)



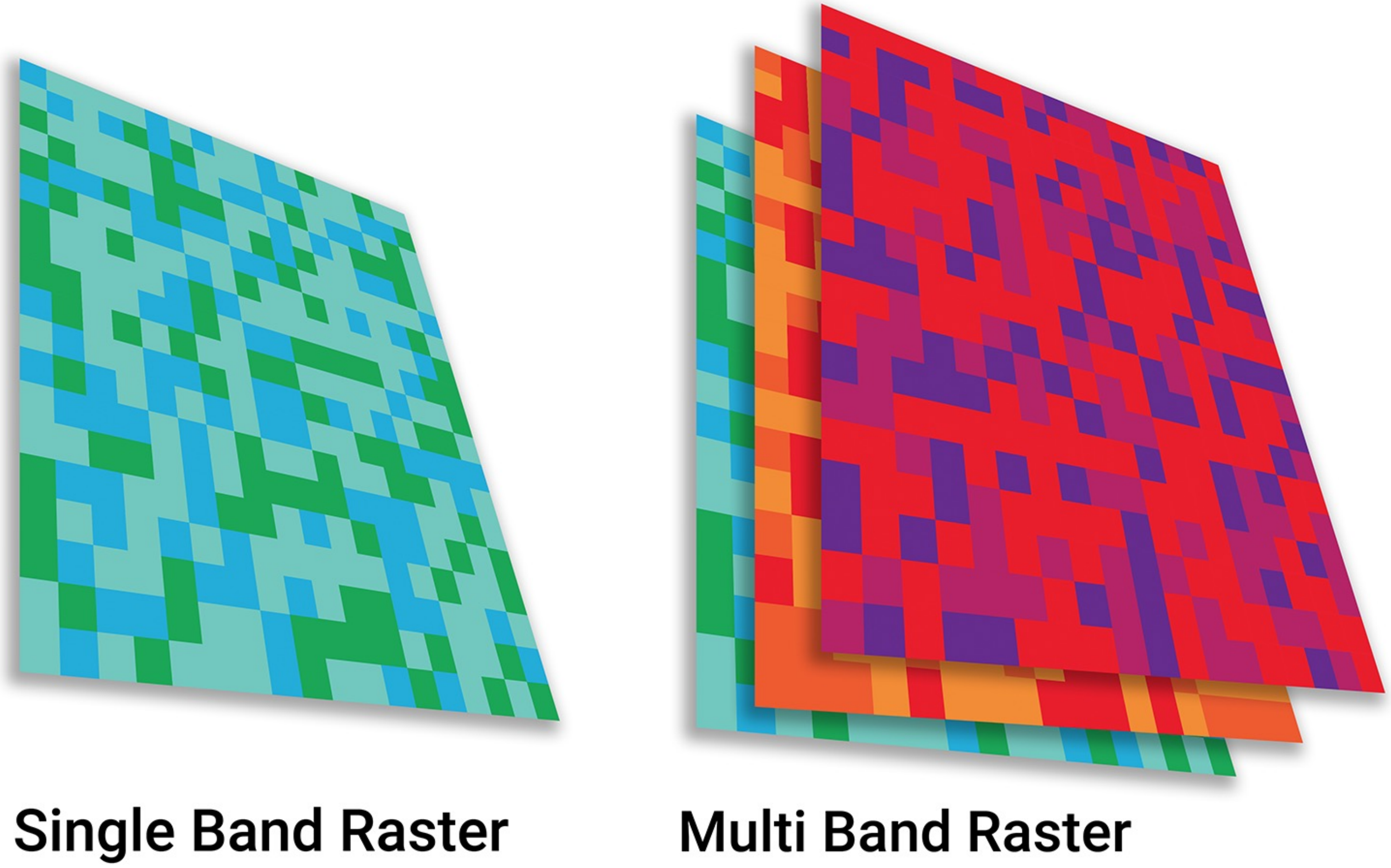
Polygons (multiple co-ordinates)



Raster (grid of values plus metadata)



Raster (grid of values plus metadata)



Raster (multiband)

Beware of projections and co-ordinate systems

Latitude, Longitude for the **whole world** in **degrees** (but which comes first?)

X, Y in projected co-ordinate systems for **specific areas** usually in **metres**

epsg.io

EPSG:27700

Share on:

OSGB36 / British National Grid -- United Kingdom Ordnance Survey

Transform

Get position on a map

Available transformations

to EPSG:4326 ▼

United Kingdom (UK) , accuracy 1.0 m, code 7710 *(default)* [grid]

United Kingdom (UK) , accuracy 2.0 m, code 1314 [7]

United Kingdom (UK) , accuracy 21.0 m, code 1195 [3]

United Kingdom (UK) - England onshore. , accuracy 10.0 m, code 1196 [3]

United Kingdom (UK) , accuracy 21.0 m, code 1197 [3]

United Kingdom (UK) , accuracy 18.0 m, code 1198 [3]

United Kingdom (UK) - Wales onshore. , accuracy 35.0 m, code 1199 [3]

United Kingdom (UK) , accuracy 3.0 m, code 5622 [3]

United Kingdom (UK) , accuracy 1.0 m, code 5339 [grid]

Selected transformation

Method: NTv2

Grid Files: OSTN15_NTv2_OSGBtoETRS.gsb

Remarks: Parameter values taken from OSGB36 to ETRS89 (2) (tfm code 7709) assuming that ETRS89 is coincident with WGS 84 within the accuracy of the tfm. Replaces OSGB36 to WGS 84 (7) (tfm code 5339).

Information source: EPSG

Revision date: 2021-08-19

Covered area powered by MapTiler

Center coordinates
305371.94 610575.38

Projected bounds:
-103976.3 -16703.87
652897.98 1199851.44

WGS84 bounds:
-9.0 49.75
2.01 61.01

United Kingdom (UK) - offshore to boundary of UKCS within 49°45'N to 61°N and 9°W to 2°E; onshore Great Britain (England, Wales and Scotland). Isle of Man onshore.

Attributes

Unit: metre

Geodetic CRS: OSGB36

Datum: Ordnance Survey of Great Britain 1936

Ellipsoid: Airy 1830

Prime meridian: Greenwich

Data source: EPSG

Information source: Ordnance Survey of Great Britain.

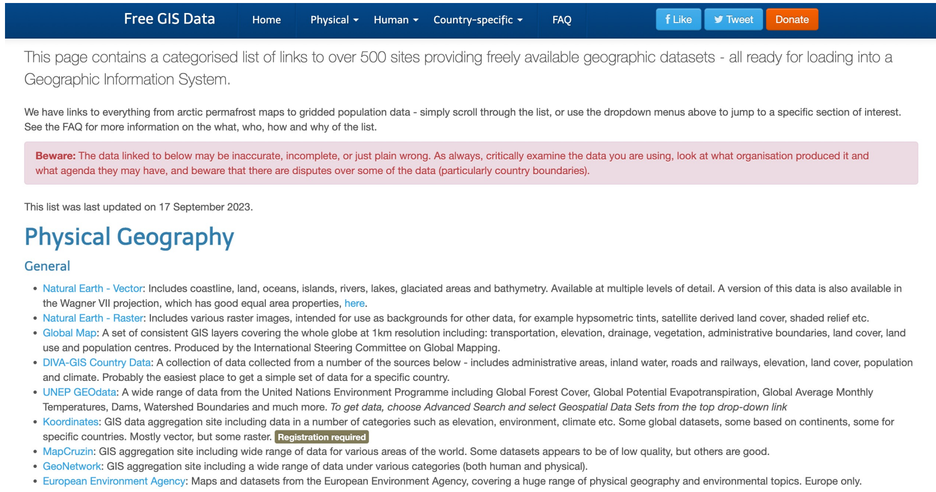
Scope: Engineering survey, topographic mapping.

Area of use: United Kingdom (UK) - offshore to boundary of UKCS within 49°45'N to 61°N and 9°W to 2°E; onshore Great Britain (England, Wales and Scotland). Isle of Man onshore.

Coordinate system: Cartesian 2D CS. Axes: easting, northing (E,N). Orientations: east, north. UoM: m.

Where to get geospatial data?

- National mapping agencies (eg. Ordnance Survey)
- National statistical organisations (eg. Office for National Statistics)
- OpenStreetMap
- Satellite operators (eg. NASA, ESA, commercial operators)
- My FreeGISData site – www.freegisdata.rtwilson.com
- **Customers, Clients…**



Tools for processing geospatial data

Command-line tools

```
Translate formats
Reproject
Resample
Get information
Query
```

Formats

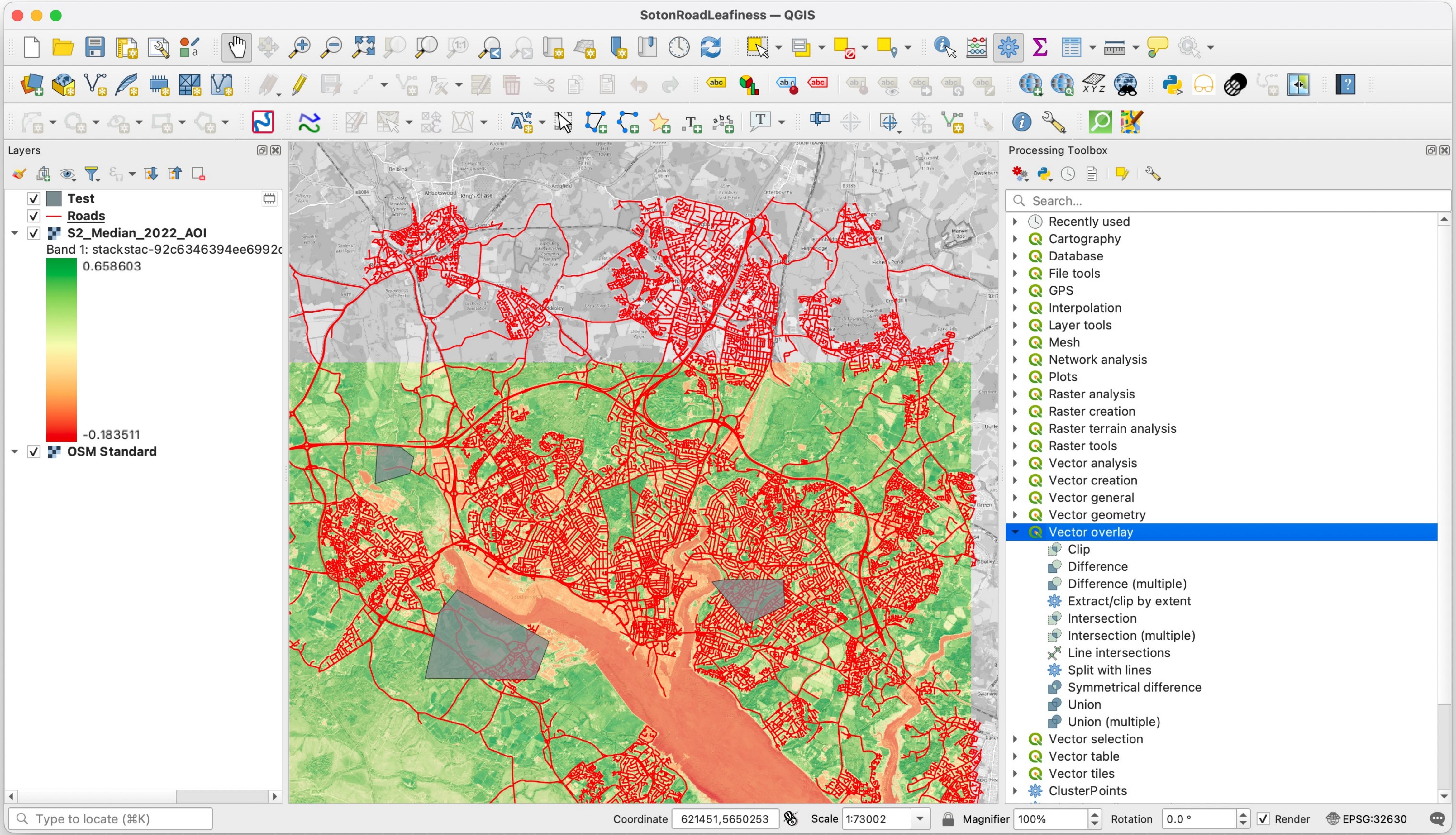
Raster	Vector
GeoTIFF / COG (.tif)	GeoJSON (.geojson)
ERDAS Imagine (.img)	GeoPackage (.gpkg)
JPEG2000 (.jp2)	GeoParquet (.parquet)
Grid (.grd or others)	Shapefile (.shp, .dbf, .shx etc)
	CSV (.csv)
	ESRI File GeoDatabase (.gdb)
	Database connections (PostGIS, Spatialite)

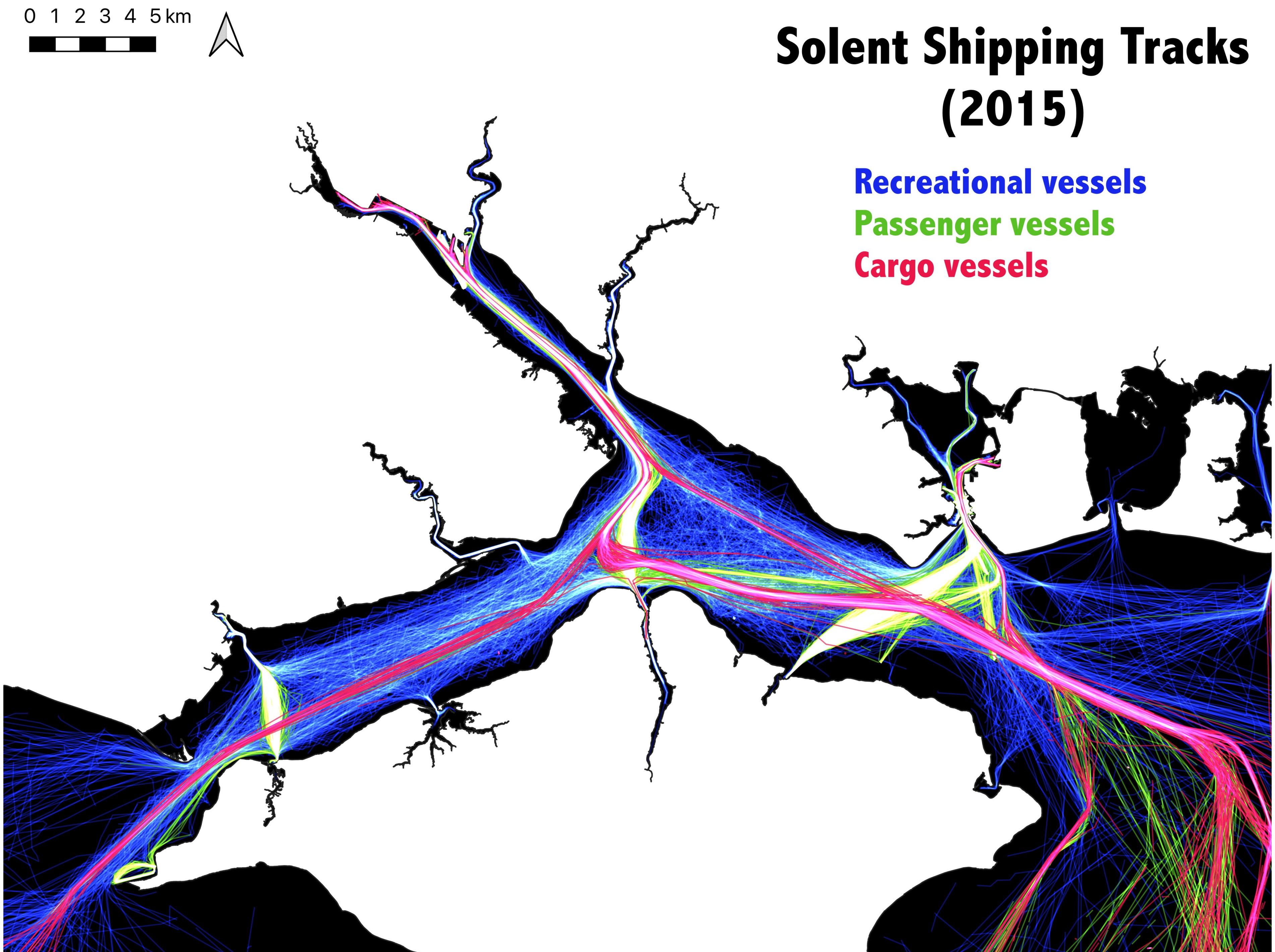
QGIS

View – Edit – Process

PyQGIS & Graphical Modeler

- Cartography
- Database
- File tools
- GPS
- Interpolation
- Layer tools
- Mesh
- Network analysis
- Plots
- Raster analysis
- Raster creation
- Raster terrain analysis
- Raster tools
- Vector analysis
- Vector creation
- Vector general
- Vector geometry
- Vector overlay
- Vector selection
- Vector table
- Vector tiles
- ClusterPoints
- Dissolve adjacent polygons
- GDAL
- Geo Simplification
- GRASS
- QuickOSM


SotonRoadLeafiness — QGIS
Layers
Test
Roads
S2_Median_2022_AOI
Band 1: stackstac-92c6346394ee6992c
0.658603
-0.183511
OSM Standard
Processing Toolbox
Search...
Recently used
Cartography
Database
File tools
GPS
Interpolation
Layer tools
Mesh
Network analysis
Plots
Raster analysis
Raster creation
Raster terrain analysis
Raster tools
Vector analysis
Vector creation
Vector general
Vector geometry
Vector overlay
Clip
Difference
Difference (multiple)
Extract/clip by extent
Intersection
Intersection (multiple)
Line intersections
Split with lines
Symmetrical difference
Union
Union (multiple)
Vector selection
Vector table
Vector tiles
ClusterPoints
Type to locate (⌘K)
Coordinate
621451,5650253
Scale
1:73002
Magnifier
100%
Rotation
0.0 °
Render
EPSG:32630


0 1 2 3 4 5 km
Solent Shipping Tracks
(2015)
Recreational vessels
Passenger vessels
Cargo vessels

PostGIS
Spatial PostgreSQL

Spatialite

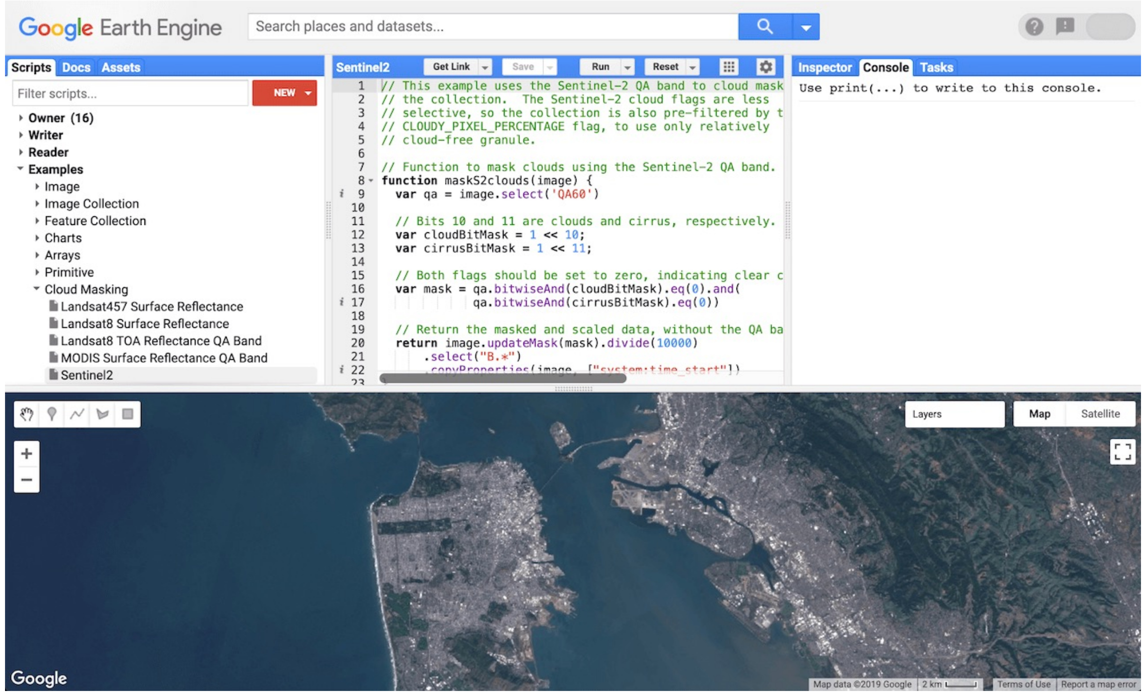
```sql
SELECT us_counties.geom , us_counties.id, COUNT(quakes.geom) AS total
FROM us_counties
JOIN quakes
  ON ST_CONTAINS(us_counties.geom, quakes.geom)
GROUP BY us_counties.id;
```

Contains / Intersects / DWithin / …

DuckDB

Cloud Native Geospatial

Google Earth Engine



STAC

COG
CLOUD OPTIMIZED
GEOTIFF

GeoParquet

Geospatial data in Parquet

Python libraries

GeoPandas

xarray

rasterstats

folium

fiona

rasterio

shapely

NumPy SciPy matplotlib

GDAL

and lots more...

What is the relationship between house prices and distance to schools?

X & Y

House Prices

	ID	Price	Date	Postcode
0	{DBA933F9-D5BC-669D-E053-6B04A8C0AD56}	205000	2022-02-18 00:00	DL9 4RS
1	{DBA933F9-D5BE-669D-E053-6B04A8C0AD56}	220000	2022-02-14 00:00	YO12 7ND
2	{DBA933F9-D5C0-669D-E053-6B04A8C0AD56}	775000	2022-02-22 00:00	HG5 0TT
3	{DBA933F9-D5C6-669D-E053-6B04A8C0AD56}	450000	2022-03-04 00:00	YO31 1BU
4	{DBA933F9-D5CC-669D-E053-6B04A8C0AD56}	175000	2022-02-25 00:00	LA2 7EB
...	...	...	...	...
912332	{F16F63C5-D419-0491-E053-6C04A8C032ED}	150000	2022-01-17 00:00	NG31 9RH
912333	{F16F63C5-D41A-0491-E053-6C04A8C032ED}	320000	2022-01-28 00:00	PE9 3FJ
912334	{F16F63C5-D41B-0491-E053-6C04A8C032ED}	340000	2022-09-01 00:00	PE23 5RS
912335	{F16F63C5-D41C-0491-E053-6C04A8C032ED}	235000	2022-03-21 00:00	LN2 4ZJ
912336	{F16F63C5-D41D-0491-E053-6C04A8C032ED}	350000	2022-04-01 00:00	NG32 2JY

Postcodes

	postcode	easting	northing
2123582	SO1 0AA	442004	111235
2123583	SO1 0AB	441967	111272
2123584	SO1 0AD	441909	111317
2123585	SO1 0AE	441864	111334
2123586	SO1 0AF	441853	111406
...	...	...	...
2164040	SO97 4AT	444918	116755
2164041	SO97 4AU	444918	116755
2164042	SO97 4AW	444918	116755
2164043	SO97 4AX	444918	116755
2164044	SO97 4AY	444918	116755

```
geometries = [shapely.geometry.Point(x, y)
              for x, y
              in zip(
                  postcodes.easting,
                  postcodes.northing)],
```

shapely

(Python vector geospatial objects)

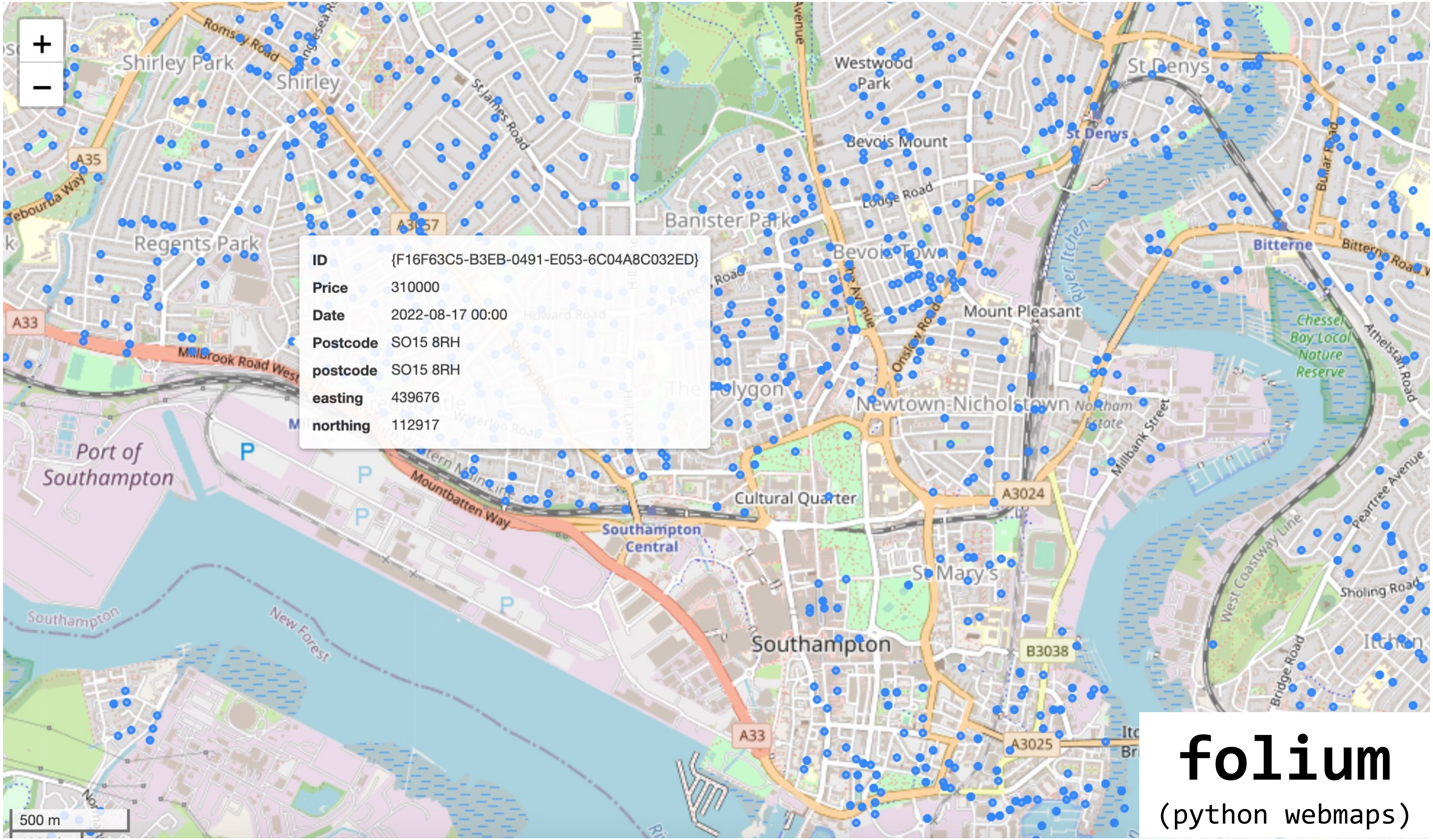
```
postcodes = gpd.GeoDataFrame(
    data=postcodes,
    geometry=geometries,
    crs=27700)
```

geopandas

(Pandas + Geo)

Index	Col1	Col2	Geometry
0	a	A	POINT(5, 9)
1	b	B	POINT(2, -3)
2	c	C	POINT(10, 7)

```python
houseprices = pd.merge(
                    left=houseprices,
                    right=postcodes,
                    left_on='Postcode',
                    right_on='postcode',
                    how='inner'
               )

houseprices.explore()
```



folium
(python webmaps)

Schools

	URN	EstablishmentName	PhaseOfEducation (name)	Easting	Northing
0	100000	The Aldgate School	Primary	533498.0	181201.0
8	100008	Argyle Primary School	Primary	530238.0	182761.0
9	100009	West Hampstead Primary School	Primary	524888.0	185067.0
10	100010	Brecknock Primary School	Primary	529912.0	184835.0
11	100011	Brookfield Primary School	Primary	528706.0	186594.0
...	...	...	...	...	...
47950	150361	Heycroft Primary School	Primary	584687.0	189038.0
47953	150365	Field Place Infant School	Primary	511692.0	103634.0
47954	150366	Heckmondwike Primary School	Primary	422235.0	423602.0
47955	150367	Mallard Primary School	Primary	455577.0	400539.0
47956	150368	Cranleigh Church of England Primary School	Primary	505646.0	139372.0

Do the same as before…

```python
joined = gpd.sjoin_nearest(houseprices,
                           schools,
                           distance_col='dist')
```

	Price	Postcode	dist
9410	320000	SO18 4JU	420.448570
10346	460000	SO40 3QS	794.062970
2965	115000	SO15 3SJ	396.339501
2562	185000	SO15 2NH	264.877330
1009	240000	SO19 9EH	554.582726
...	...	...	...
9740	481000	SO15 7QZ	497.800161
5495	270000	SO41 8YD	891.809957
1575	352000	SO19 8GB	601.658541
3780	371000	SO30 2FF	512.621693
7560	599950	SO41 0LL	1454.759430

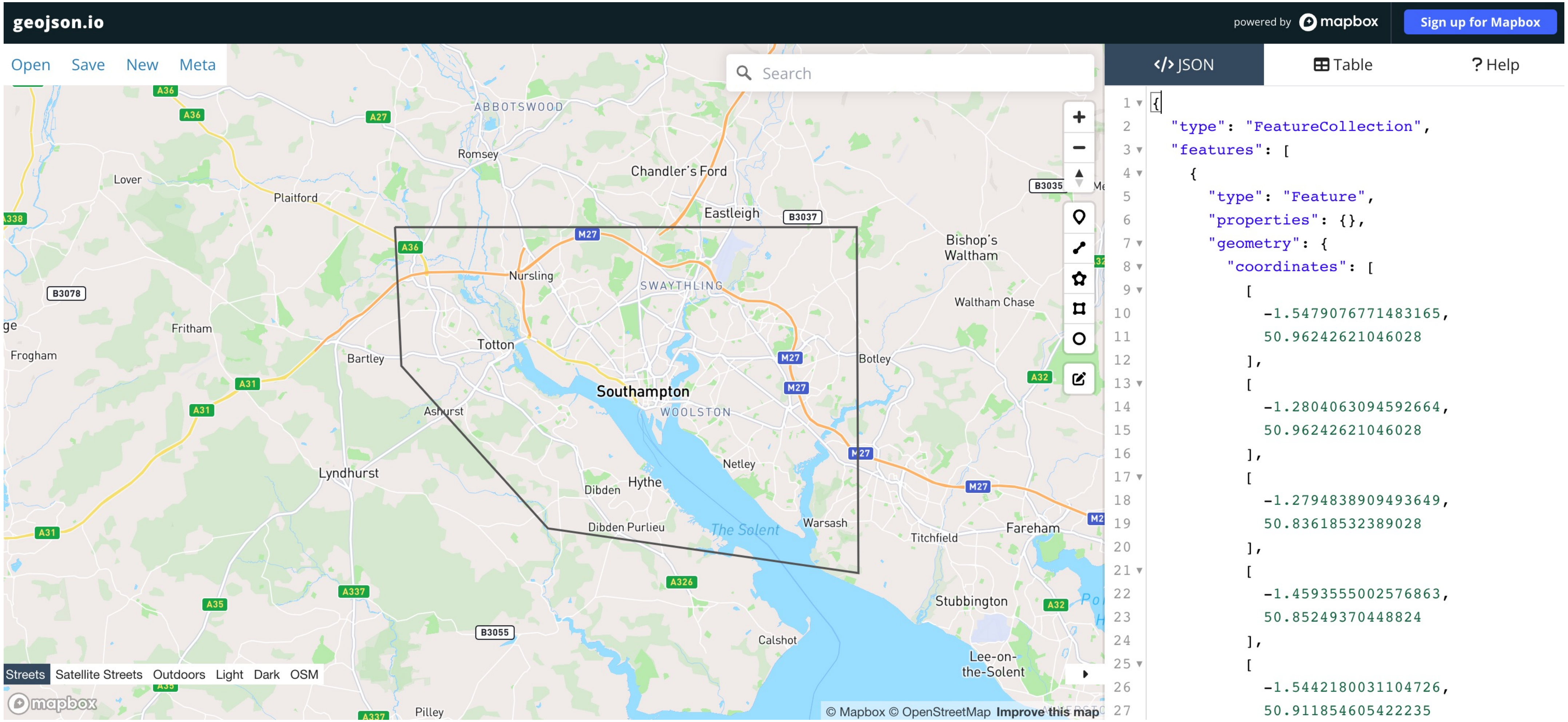
```
count    14596.000000
mean       586.641433
std        453.603914
min          0.000000
25%        309.499999
50%        486.834146
75%        706.510439
max       5307.664835
Name: dist, dtype: float64
```

How leafy are different neighbourhoods of Southampton?


Spindlewood Cl


Northumberland

```python
# Create the Dask cluster
cluster = GatewayCluster()
client = cluster.get_client()

cluster.adapt(minimum=4, maximum=24)
print(cluster.dashboard_link)
```




```
area_of_interest = <GEOJSON>

bbox = rasterio.features.bounds(area_of_interest)
```

rasterio

(python raster I/O)

```python
stac = pystac_client.Client.open(
      "https://planetarycomputer.microsoft.com/api/stac/v1",
      modifier=planetary_computer.sign_inplace,
)

search = stac.search(
      bbox=bbox,
      datetime="2022-01-01/2023-01-01",
      collections=["sentinel-2-l2a"],
      query={"eo:cloud_cover": {"lt": 10}},
)

items = search.item_collection()
```

Microsoft Planetary Computer

STAC

COG CLOUD OPTIMIZED GEOTIFF

37 items

(individual satellite images from different times)

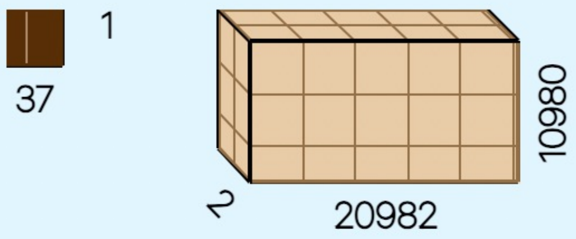
pystac

(python STAC interface)

Near Infra-red and Red bands

```
data = stackstac.stack(
            items,
            assets=["B08", "B04"],
            chunksize=4096,
            resolution=10,
)
```

	Array	Chunk
Bytes	127.02 GiB	128.00 MiB
Shape	(37, 2, 10980, 20982)	(1, 1, 4096, 4096)
Dask graph	1332 chunks in 5 graph layers	
Data type	float64 numpy.ndarray	

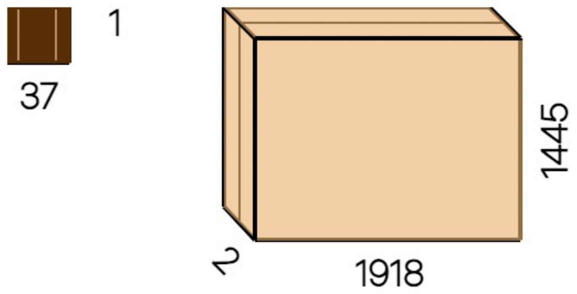


xarray & stackstac

(labelled multi-dim data & combining STAC items)

```python
data = stackstac.stack(
            items,
            assets=["B08", "B04"],
            chunksize=4096,
            resolution=10,
            bounds_latlon=bbox
)
```

	Array	Chunk
Bytes	1.53 GiB	21.14 MiB
Shape	(37, 2, 1445, 1918)	(1, 1, 1445, 1918)
Dask graph	74 chunks in 5 graph layers	
Data type	float64 numpy.ndarray	

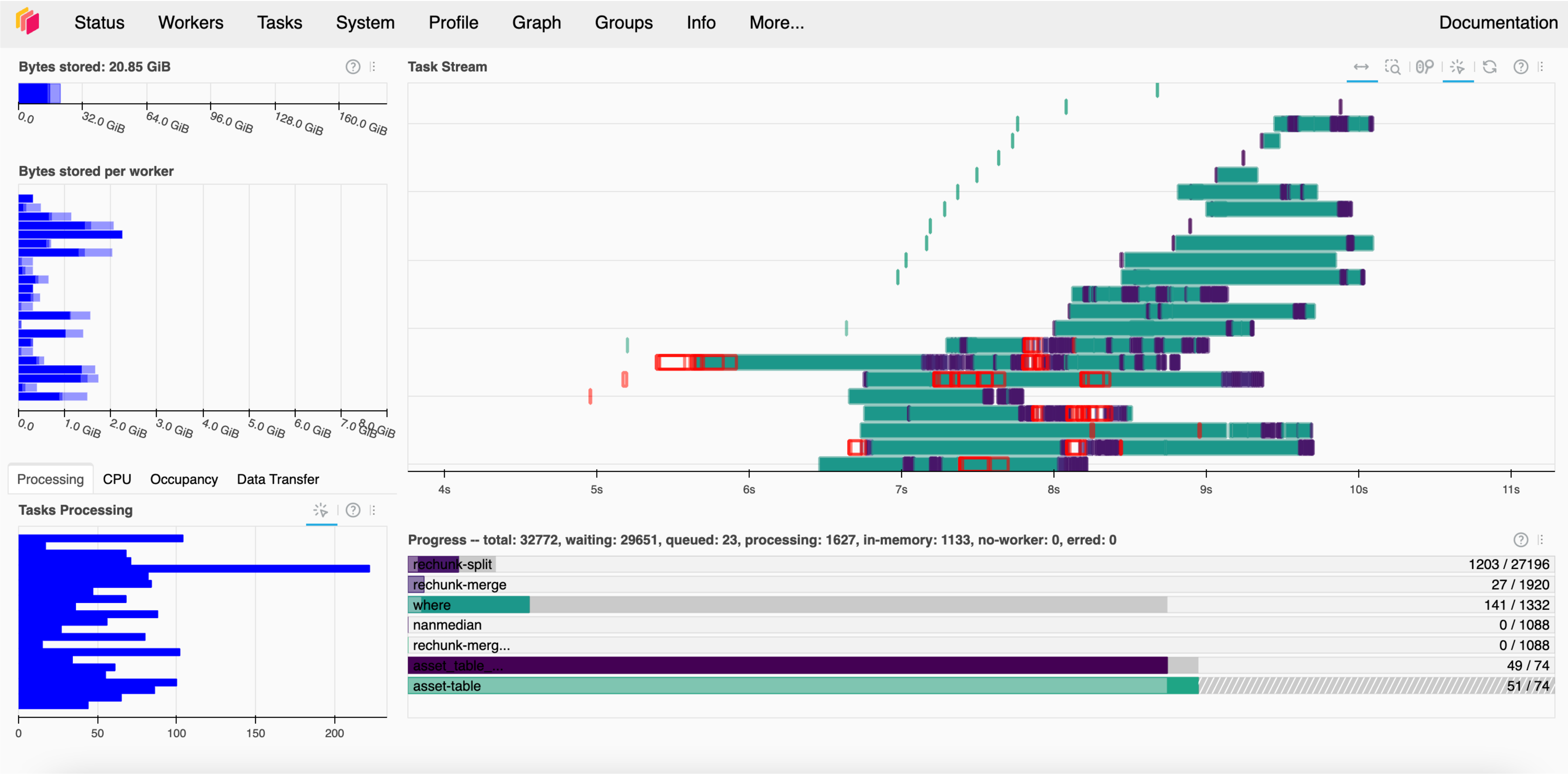


```
median = data.median(dim="time").compute()
```

Small: 15s

Large: 3 minutes

```
nir = median.sel(band='nir')
red = median.sel(band='red')

ndvi = (nir - red)/(nir + red)

ndvi.rio.to_raster("S2_Median_2022_AOI.tif")
```



Status
Workers
Tasks
System
Profile
Graph
Groups
Info
More...
Documentation
Bytes stored: 20.85 GiB
0.0
32.0 GiB
64.0 GiB
96.0 GiB
128.0 GiB
160.0 GiB
Bytes stored per worker
0.0
1.0 GiB
2.0 GiB
3.0 GiB
4.0 GiB
5.0 GiB
6.0 GiB
7.0 GiB
8.0 GiB
Processing
CPU
Occupancy
Data Transfer
Tasks Processing
0
50
100
150
200
Task Stream
4s
5s
6s
7s
8s
9s
10s
11s
Progress -- total: 32772, waiting: 29651, queued: 23, processing: 1627, in-memory: 1133, no-worker: 0, erred: 0
rechunk-split
1203 / 27196
rechunk-merge
27 / 1920
where
141 / 1332
nanmedian
0 / 1088
rechunk-merg...
0 / 1088
asset_table_...
49 / 74
asset-table
51 / 74

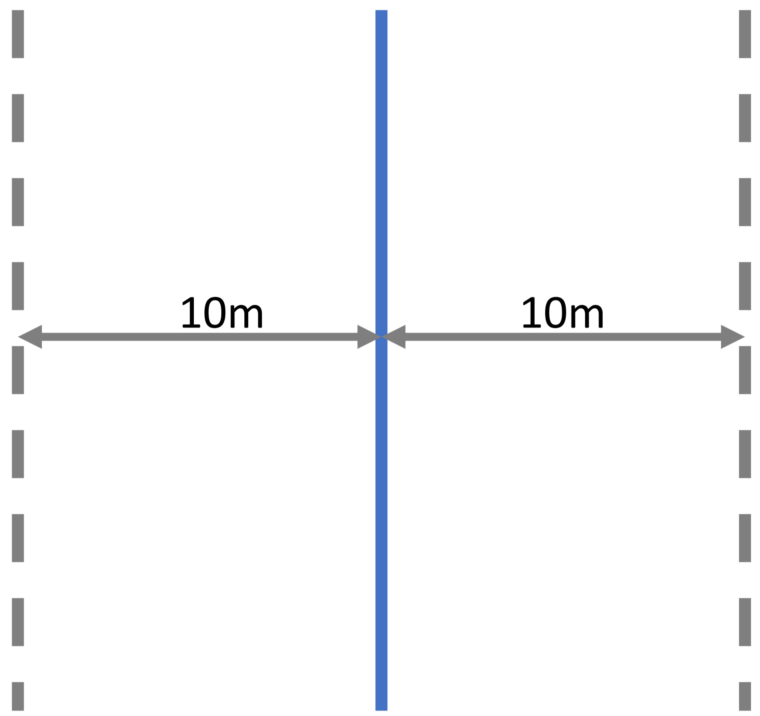

-0.18
0.66

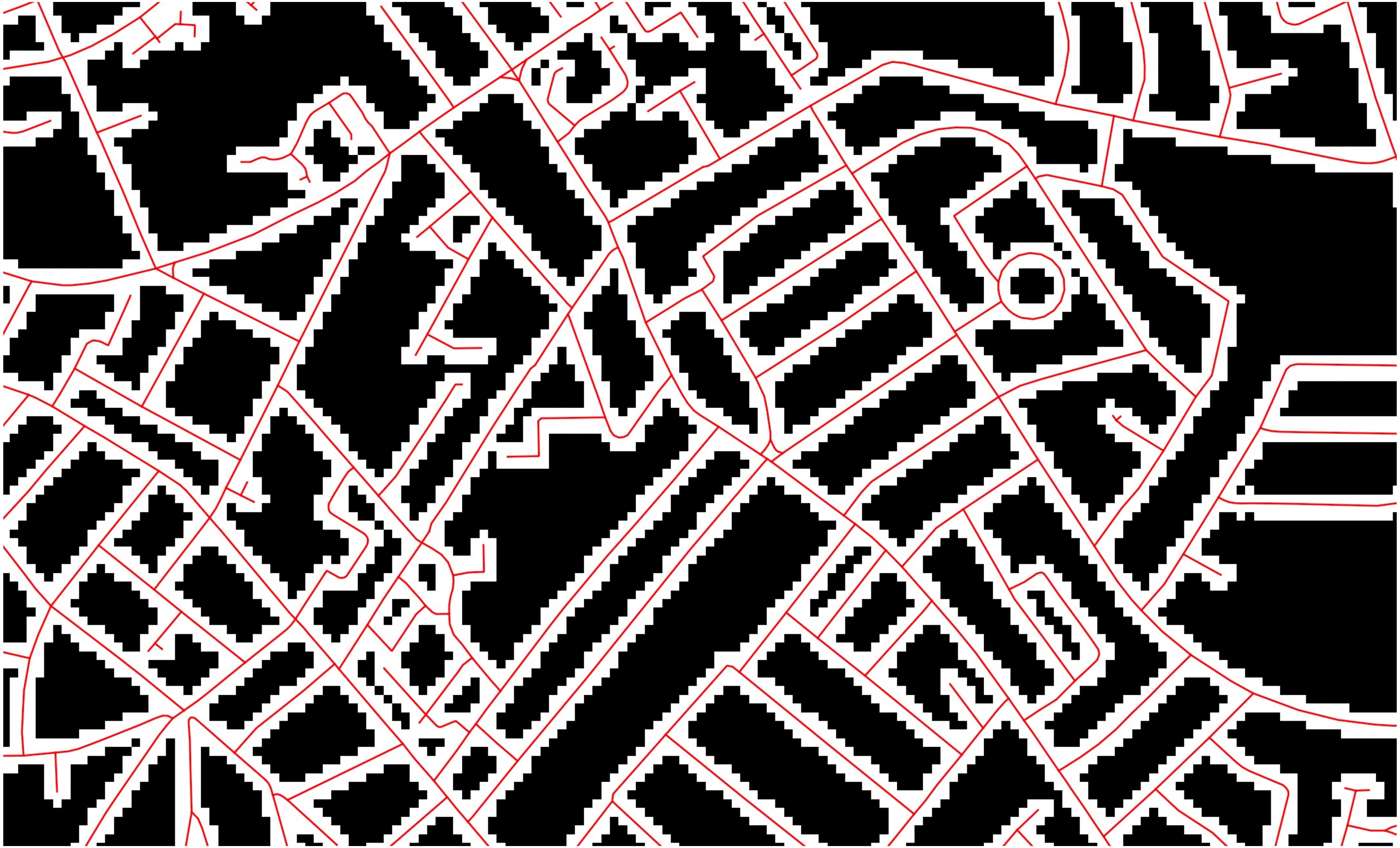
osmnx

(OSM network extraction)

```python
roads = osmnx.graph_from_point(
                        (50.91404, -1.41640),
                        dist=10000,
                        network_type="drive")

osmnx.save_graph_geopackage(roads, "SotonRoads.gpkg")
```

```
roads.geometry = roads.geometry.buffer(10)
```

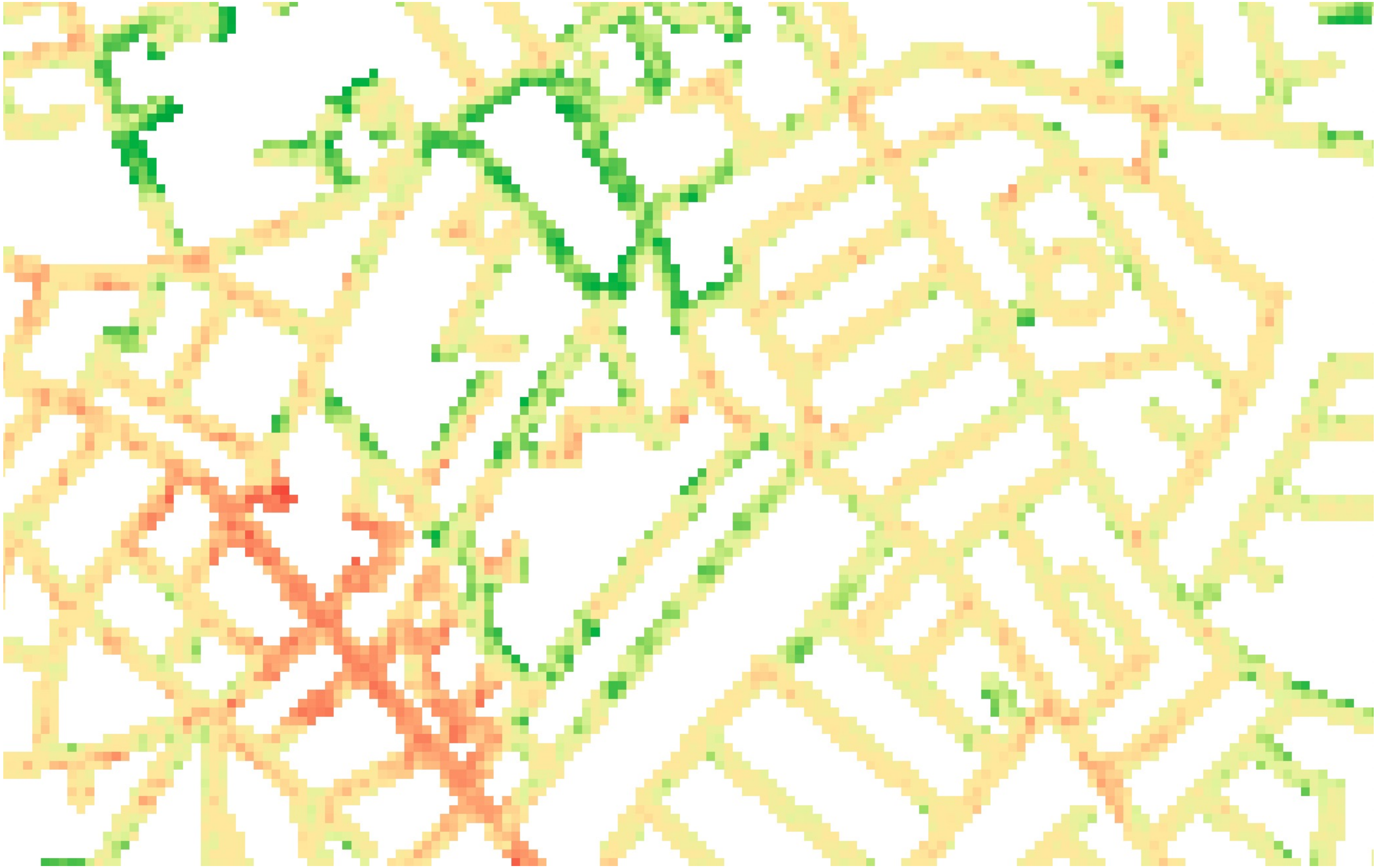




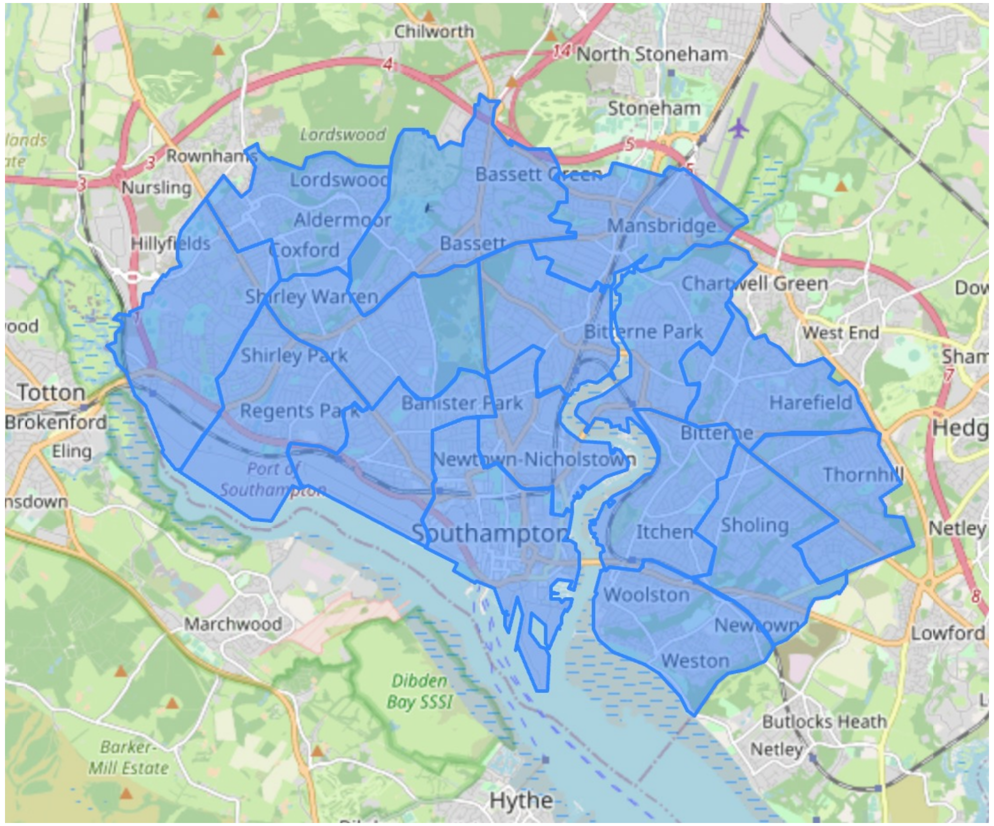
```python
with rio.open("AOI_Roads.tif") as f:
     roads_raster = f.read(1)
     meta = f.meta

with rio.open("S2_Median_2022_AOI.tif") as f:
     ndvi_raster = f.read(1)

masked = roads_raster * ndvi_raster

with rio.open("Masked_NDVI.tif", 'w', **meta) as f:
     f.write(masked, 1)
```

rasterio

(python raster I/O)



```python
wards = gpd.read_file("SotonWards.shp")

wards = wards.to_crs(meta['crs'])
```



geopandas

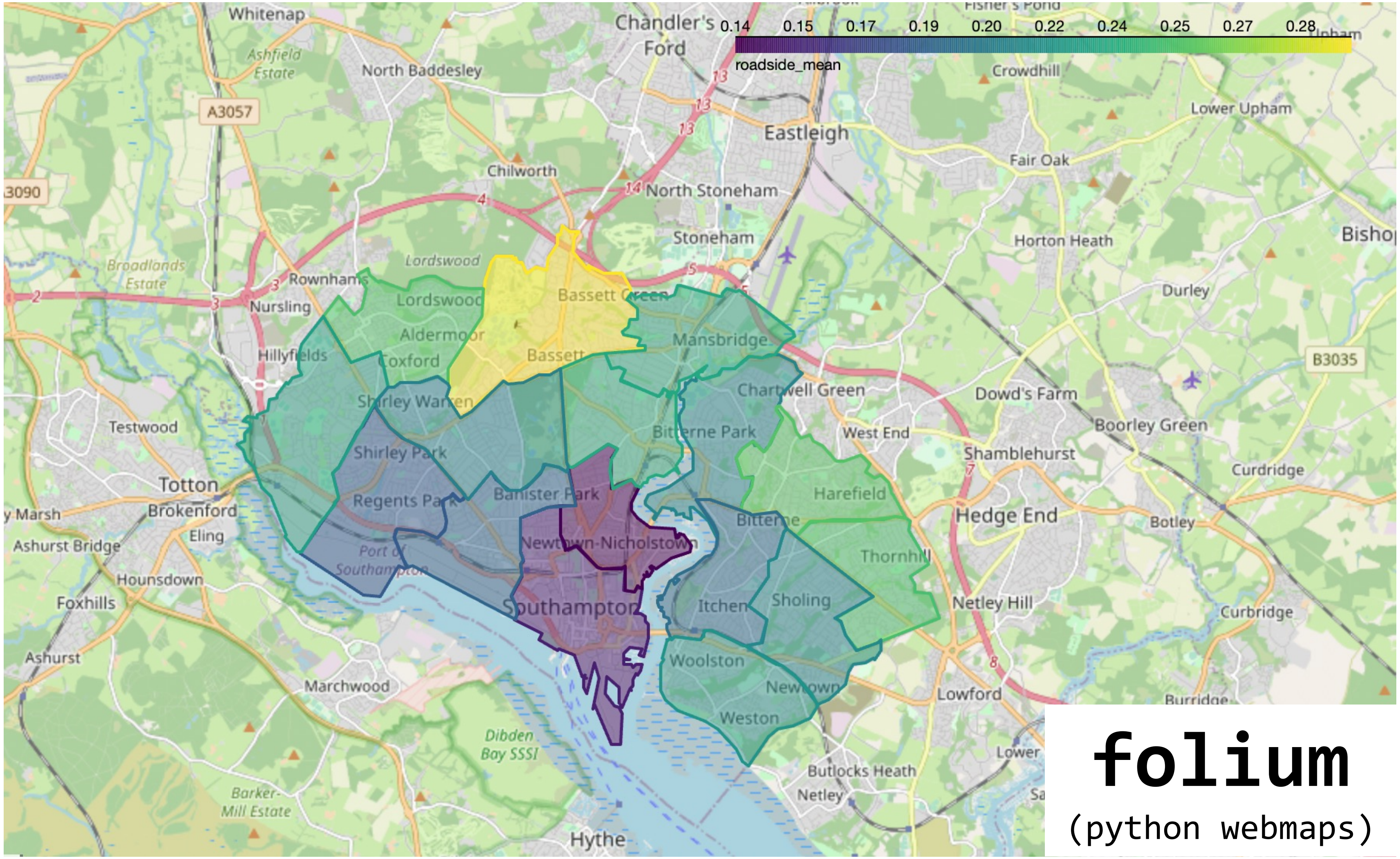
(Pandas + Geo)

```python
roadside_stats = zonal_stats(
                    vectors=wards.geometry,
                    raster='Masked_NDVI.tif',
                    stats='mean',
                    nodata=0,
                    prefix="roadside_"
                )

wards = wards.join(pd.DataFrame(roadside_stats),
                how='left')
```

rasterstats

(python zonal statistics)

```python
wards.explore("roadside_mean")
```



folium

(python webmaps)

	wd17cd	wd17nm	roadside_mean	overall_mean
0	E05002455	Bargate	0.151769	0.152316
1	E05002456	Bassett	0.297117	0.383503
2	E05002457	Bevois	0.138306	0.148619
3	E05002458	Bitterne	0.250498	0.304511
4	E05002459	Bitterne Park	0.212127	0.312092
5	E05002460	Coxford	0.252240	0.323646
6	E05002461	Freemantle	0.190042	0.179213
7	E05002462	Harefield	0.254662	0.314489
8	E05002463	Millbrook	0.183902	0.170176
9	E05002464	Peartree	0.195554	0.263064
10	E05002465	Portswood	0.235529	0.306308
11	E05002466	Redbridge	0.229053	0.264243
12	E05002467	Shirley	0.209484	0.327046
13	E05002468	Sholing	0.214263	0.300729
14	E05002469	Swaythling	0.234186	0.326982
15	E05002470	Woolston	0.220839	0.314683

…?

Any questions?

I do freelance geospatial software engineering and data science

www.rtwilson.com
robin@rtwilson.com