Intro to Spatial Measurements

How do we describe spatial patterns?

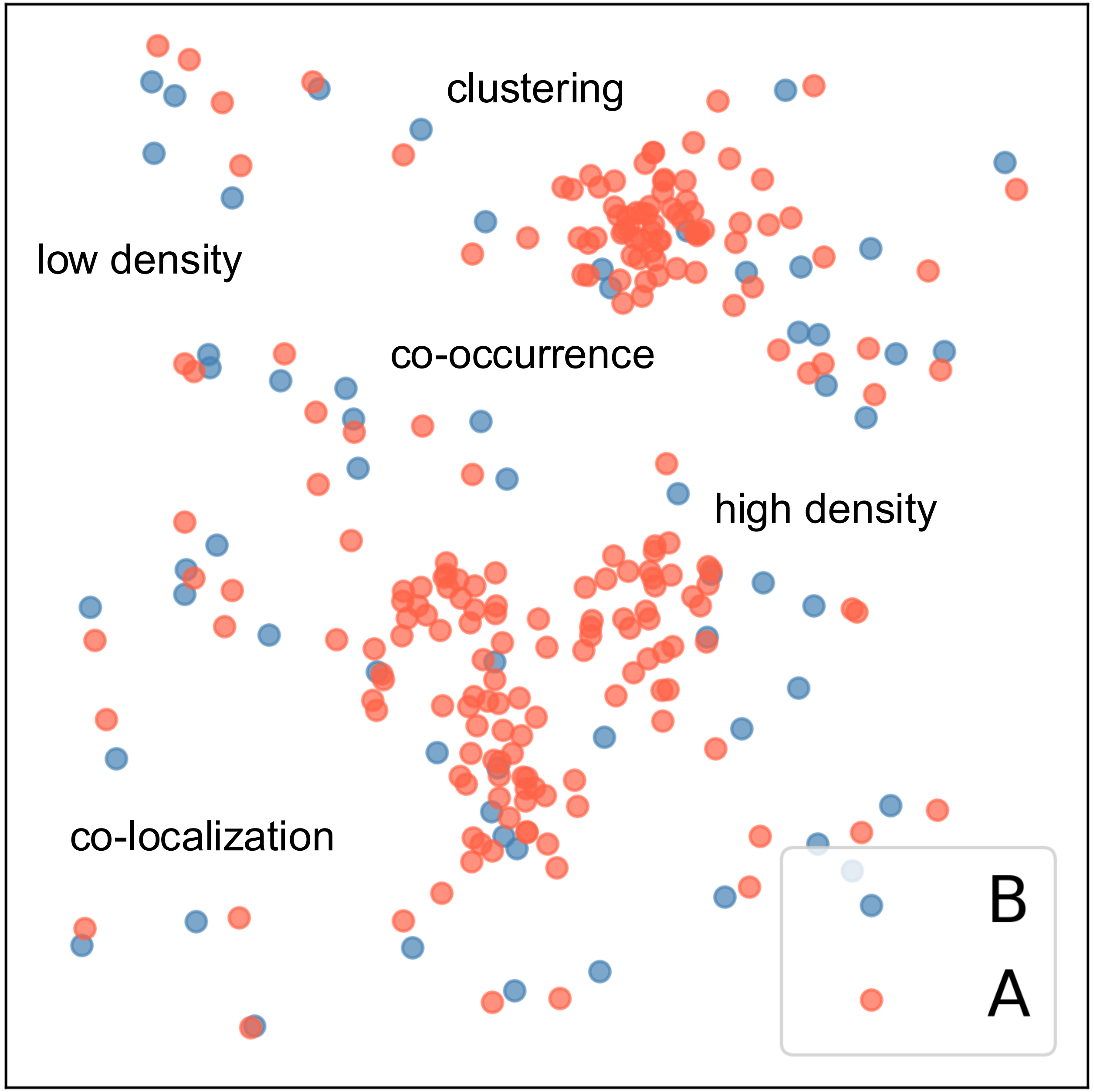
MAX BRAMBACH | OYLER-YANIV LAB

What do you see in this plot?

We have the right intuition for describing spatial patterns

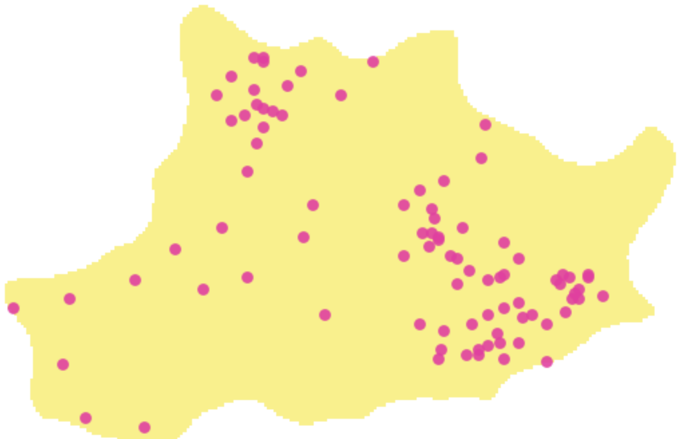
To draw conclusions from spatial patterns we need to standardize how we describe them.

For this we have to perform more computation, since the only direct measurements of spatial patterns are the coordinates of the objects.



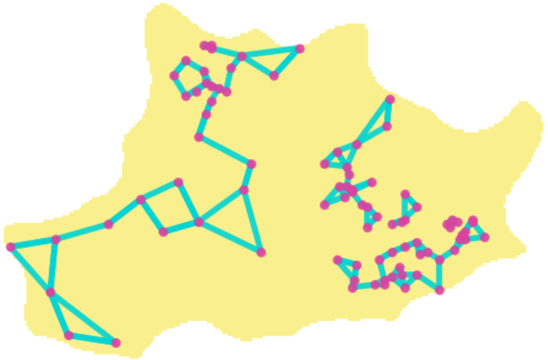
Coordinates are not data

What we see

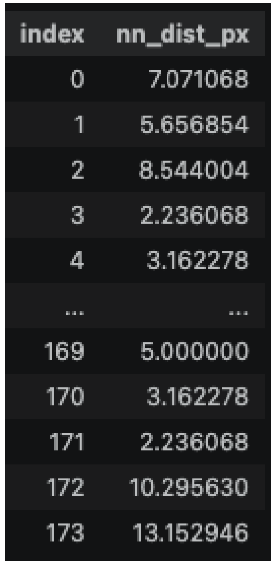


Spot distribution

What we need

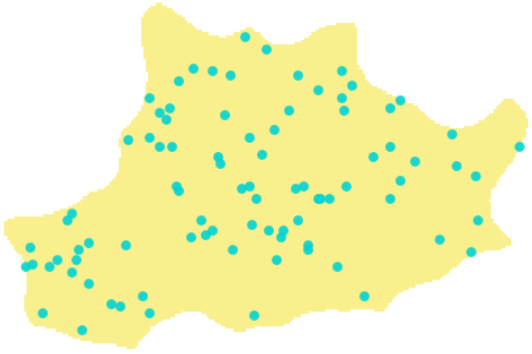


Relationships



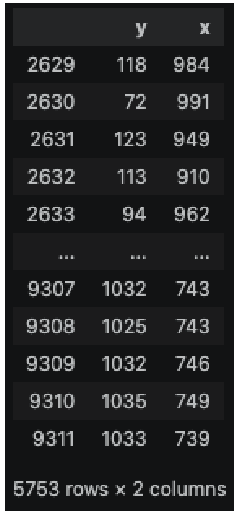
index	nn_dist_px
0	7.071068
1	5.656854
2	8.544004
3	2.236068
4	3.162278
...	...
169	5.000000
170	3.162278
171	2.236068
172	10.295630
173	13.152946

Measurements



Null distribution

What we actually have

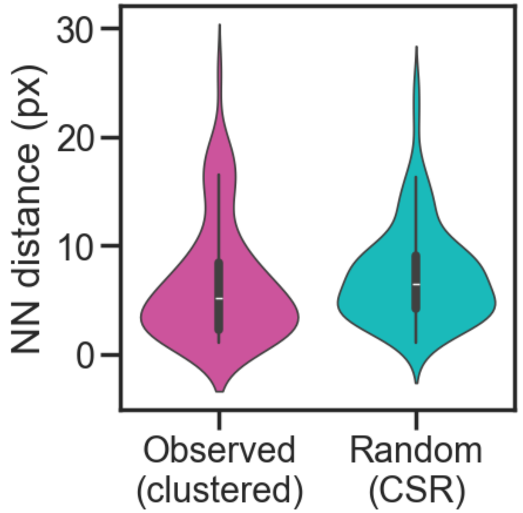


	y	x
2629	118	984
2630	72	991
2631	123	949
2632	113	910
2633	94	962
...	...	...
9307	1032	743
9308	1025	743
9309	1032	746
9310	1035	749
9311	1033	739

5753 rows × 2 columns

Coordinates

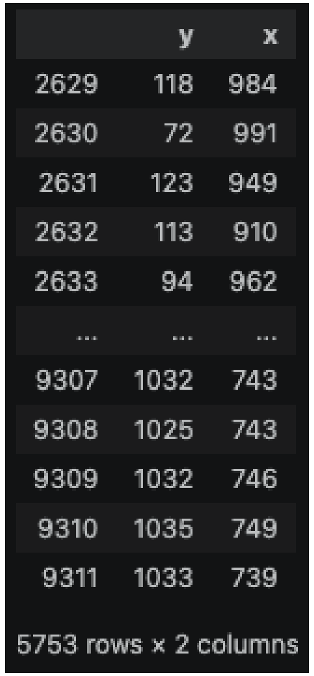
What we want



Results

From coordinates to data is a series of choices

Coordinates



	y	x
2629	118	984
2630	72	991
2631	123	949
2632	113	910
2633	94	962
...	...	...
9307	1032	743
9308	1025	743
9309	1032	746
9310	1035	749
9311	1033	739

5753 rows × 2 columns

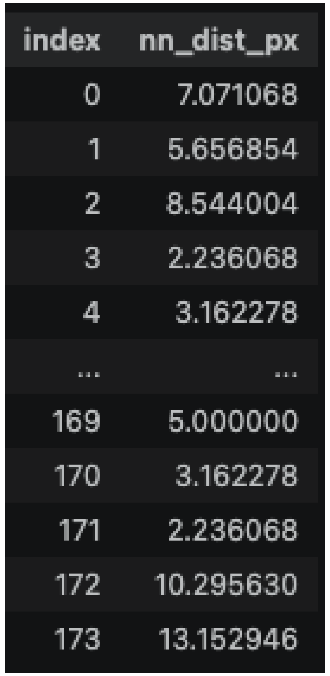
Your scientific question

<u>Create spatial relationships</u>
- nearest neighbors
- touching cells
- objects within a certain distance

Measure <u>properties</u> of those relationships
- distance to nearest neighbor
- number of objects within a distance
- types of objects within a distance

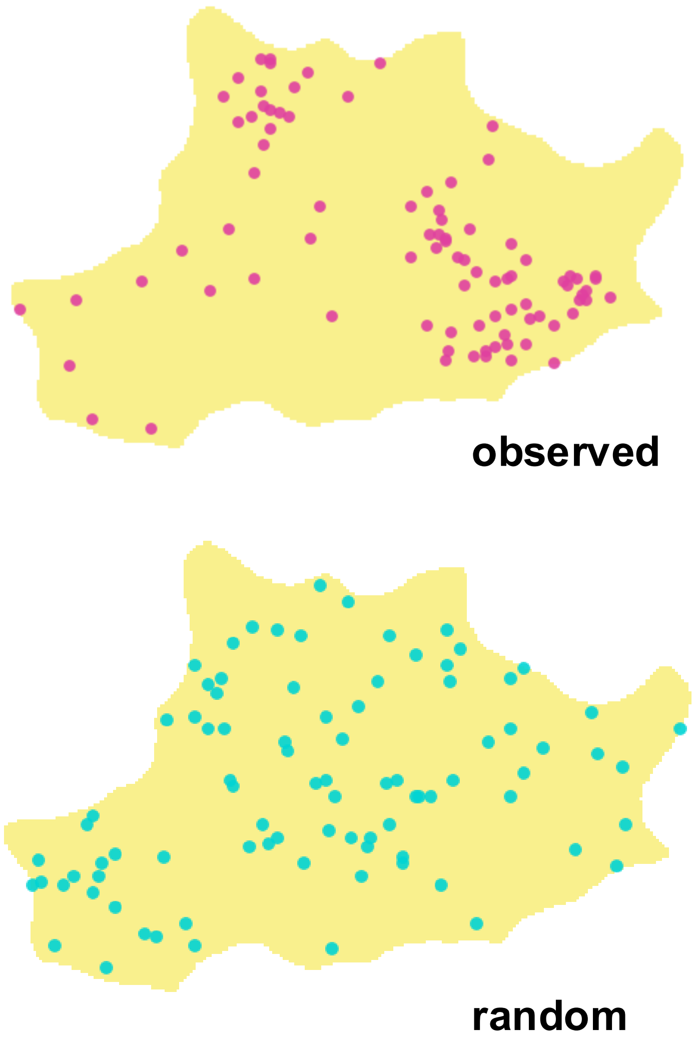
Compare against a <u>null-distribution</u>
- complete spatial randomness
- image randomization
- label-swapping

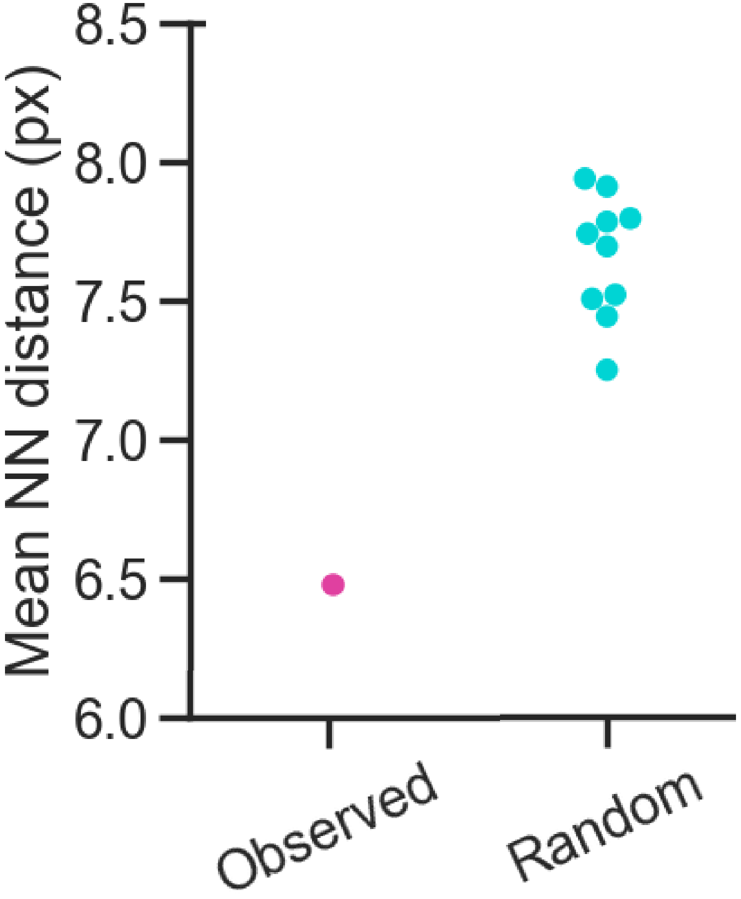
Data



index	nn_dist_px
0	7.071068
1	5.656854
2	8.544004
3	2.236068
4	3.162278
...	...
169	5.000000
170	3.162278
171	2.236068
172	10.295630
173	13.152946

Analysing spatial patterns requires a null distribution





Cells are clustered – compared to what?

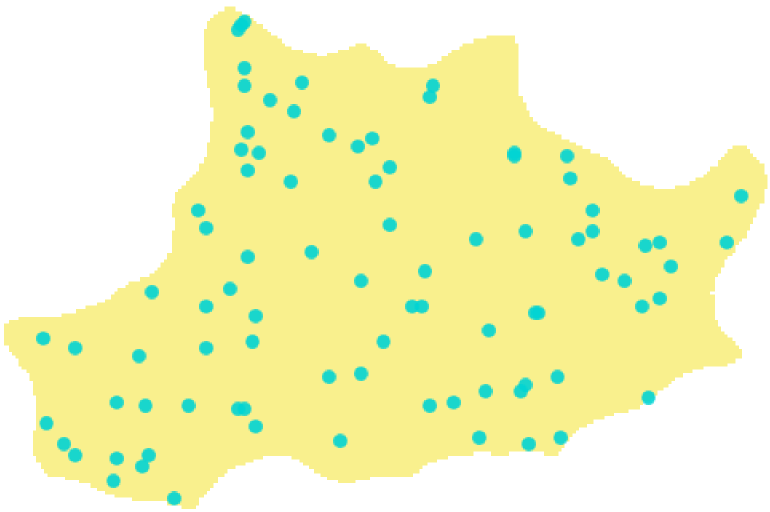
Proteins colocalize – compared to what?

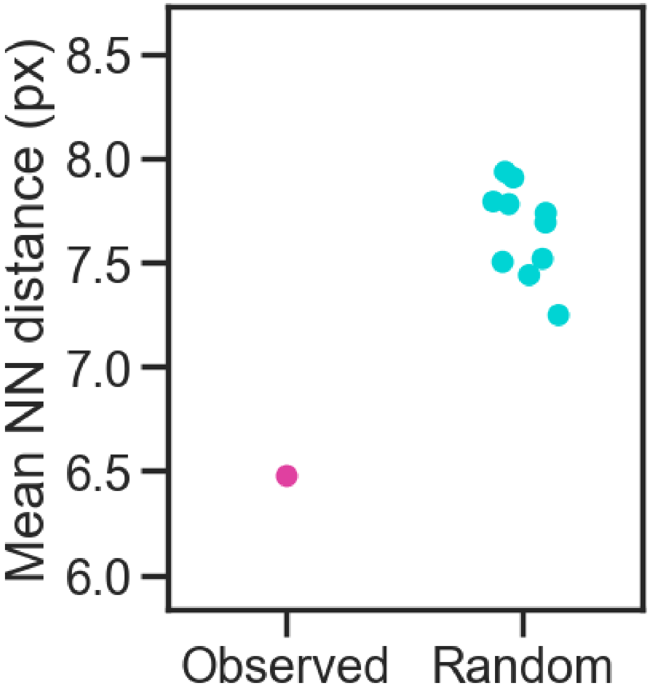
Cell density is high – compared to what?

Properties of a good null distribution:

- Reflect aspects of the observed distribution
  (e.g. same number of objects)
- Differ in meaningful aspects
  (e.g. different / random spatial distribution)

A common null distribution is complete spatial randomness Monte Carlo





Complete Spatial Randomness:

- Conserves the number of objects.
- Randomizes the position of objects.

Monte Carlo:

- One random distribution is not reliable.
- Repeatedly generate CSR distributions.
- Iterations create more stable average measurements and p-values.

Strategy:

- Randomize the positions of the observed spots.
- Only place the randomized points <u>where observed spots could have been</u> (e.g. inside the cell).

Benefits:

- Handles arbitrary geometry for free.
- Non-parametric.

Limitations:

- Can be computationally expensive
- Can be too ‘aggressive’ as a null; i.e., it destroys all spatial structure.

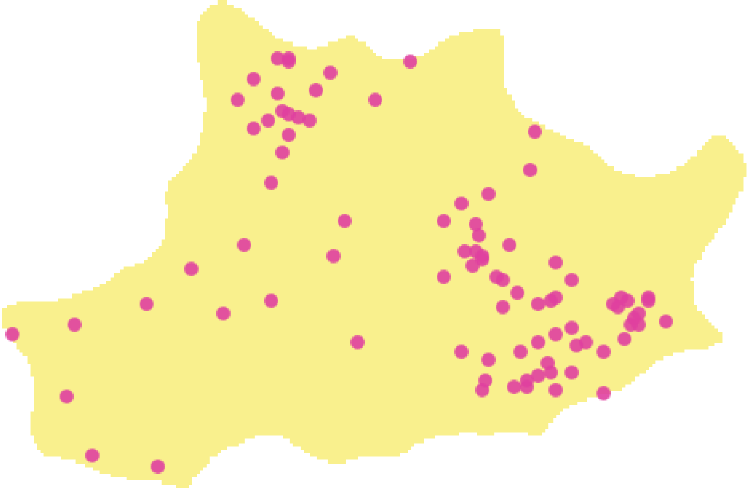
Other null distributions

Null strategy	Preserves (controls for)	Destroys (tests for)	Aggressiveness	Use when
Label randomisation	All positions of all objects	Type identity (A vs B)	Gentlest	Two populations: "do A and B associate/segregate?" Holds all spatial structure fixed, tests only the relationship between types.
Toroidal shift / translation (points or images)	Each pattern's internal structure	Relative position between patterns	Very gentle	Two patterns/channels: "is A's position related to B's?" Preserves each one's own clustering, randomises only their alignment.
Inhomogeneous / density-conditioned	Global density (e.g. a gradient)	Fine-scale arrangement	Moderate	Single population in non-uniform tissue: "is there clustering *beyond* the density gradient?" Fixes CSR's false-clustering on gradients.
Hard-core CSR	Object count + minimum spacing	Arrangement above the spacing scale	Moderate	Single population where objects have physical size: avoids CSR's false "dispersion" at short range.
CSR (complete spatial randomness)	Object count + region shape	All spatial arrangement	Aggressive	Single population, roughly uniform region: "is there *any* structure?" The default; wrong when real density varies.
Chunk / block (Costes)	Sub-block (sub-resolution) structure	Correlations above block size	Aggressive (tunable by block size)	Pixel co-localisation significance: block ≈ PSF size, so it respects that sub-resolution pixels are inherently coupled.
Pixel randomisation	Only the intensity histogram	All spatial structure	Most aggressive	Rarely ideal — destroys real autocorrelation (membranes, continuity), so almost anything looks significant. Use as a sanity floor, not a serious null.

What determines the choice of null distribution?

YOUR SCIENTIFIC QUESTION

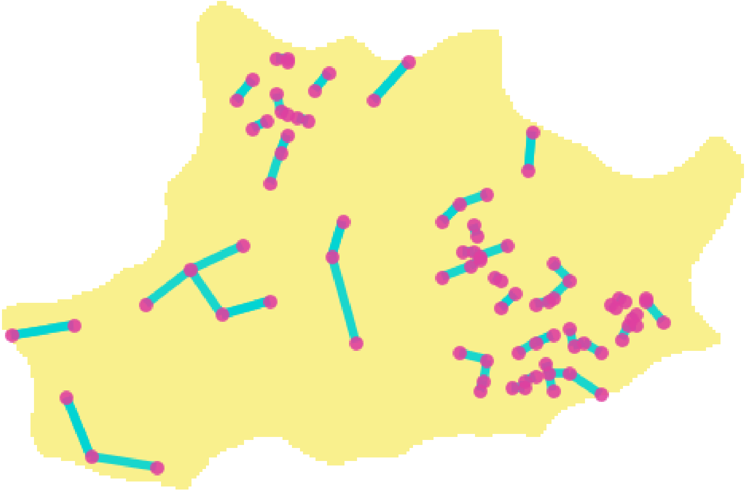
Spatial analysis requires connections between objects



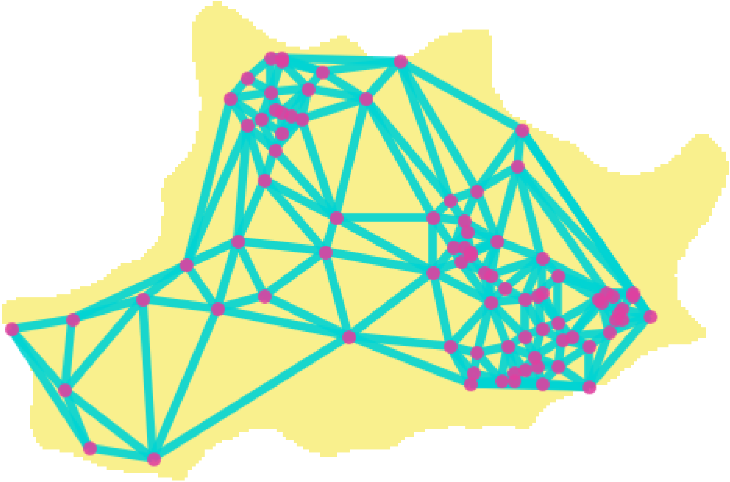
Spot distribution / raw coordinates

Define under which conditions two objects are connected

We have to work for this!



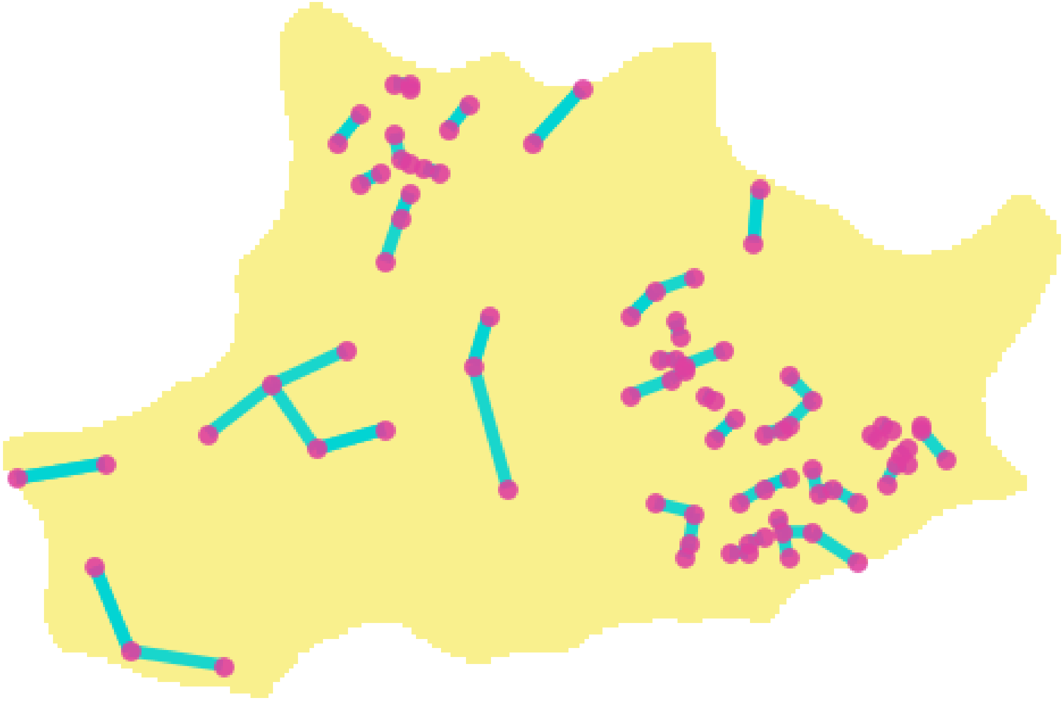
Nearest neighbor



Delaunay graph

Note: Pixel-based analysis methods (see later) use the shared pixel grid to establish spatial connections.

An effective way to establish spatial relationships is the nearest neighbour search.



Algorithm
- For each spot
- find the nearest other spot

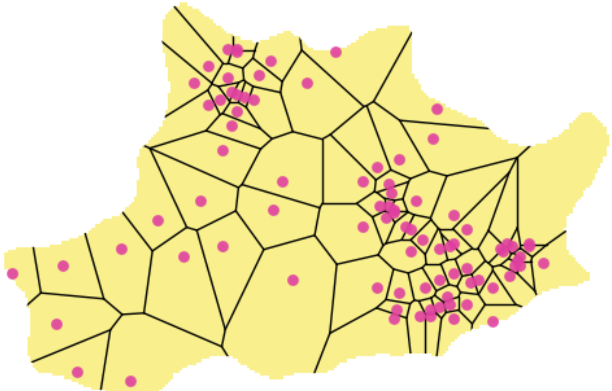
Variations
- Find the k nearest neighbors
- Radius queries: connect all spots within a specified radius

Benefits
- intuitive
- captures local desnity

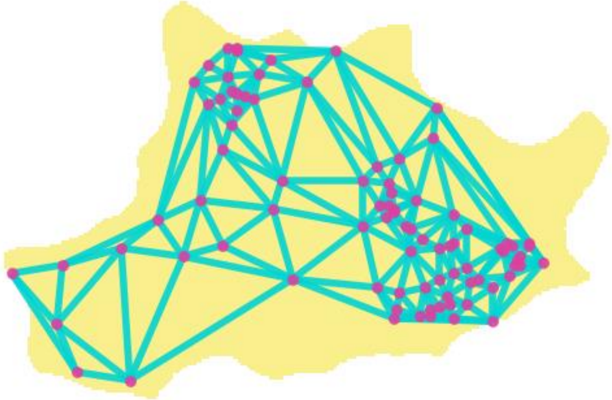
Limitations
- Can under-connect points and produce sparse graphs
- k=1 can be sensitive to micropatterns

Other formalisms to create spatial relationships exist

Spatial partitioning



Voronoi diagram

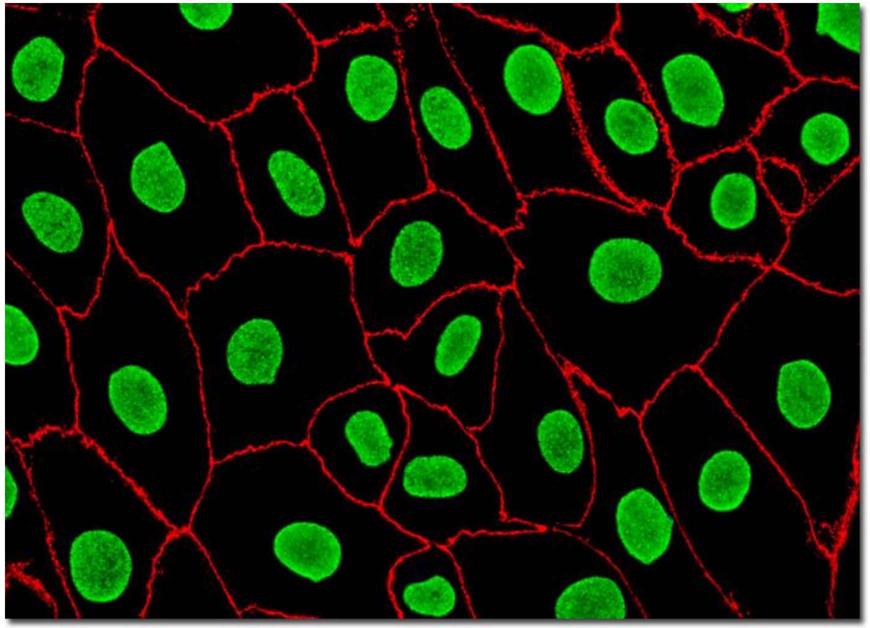


Delaunay graph

Benefits
- parameter free
- creates a hierarchy of neighbors

Limitations
- might connect unrelated objects over longer distances

Biological interaction



e.g. touching membranes

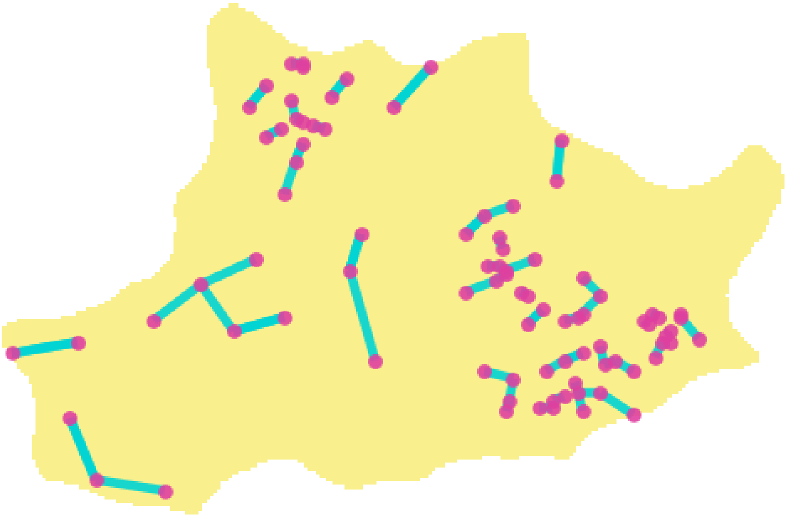
Benefits
- biologically meaningful
- creates a hierarchy of neighbors

Limitations
- often difficult to establish
- technically advanced

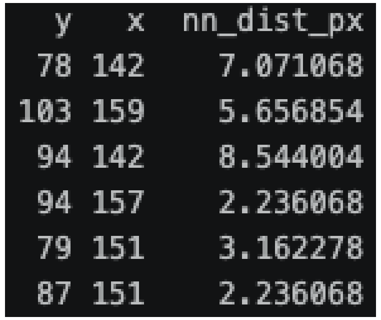
What determines how you establish spatial relationships?

YOUR SCIENTIFIC QUESTION

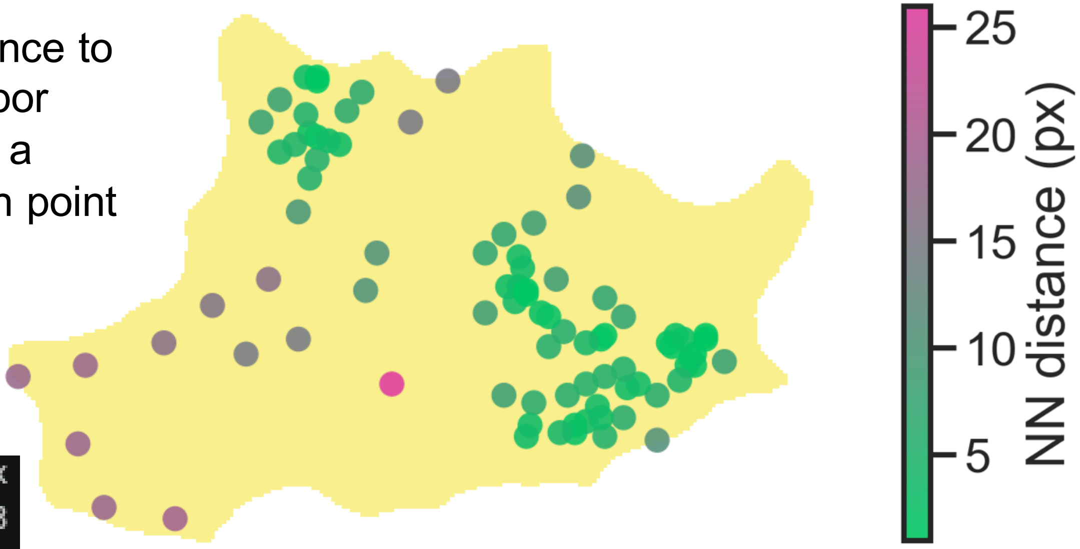
The most basic spatial feature is density / distance to neighbors



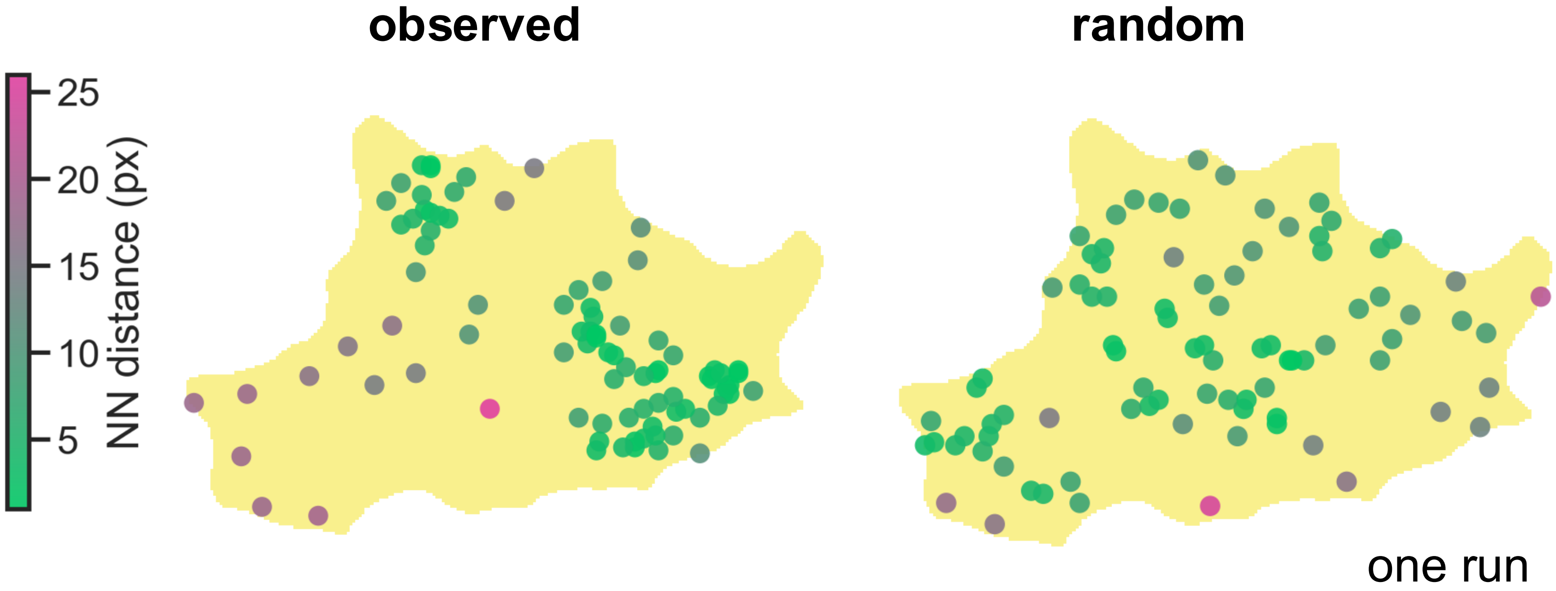
Measure distance to nearest neighbor and store it as a feature of each point



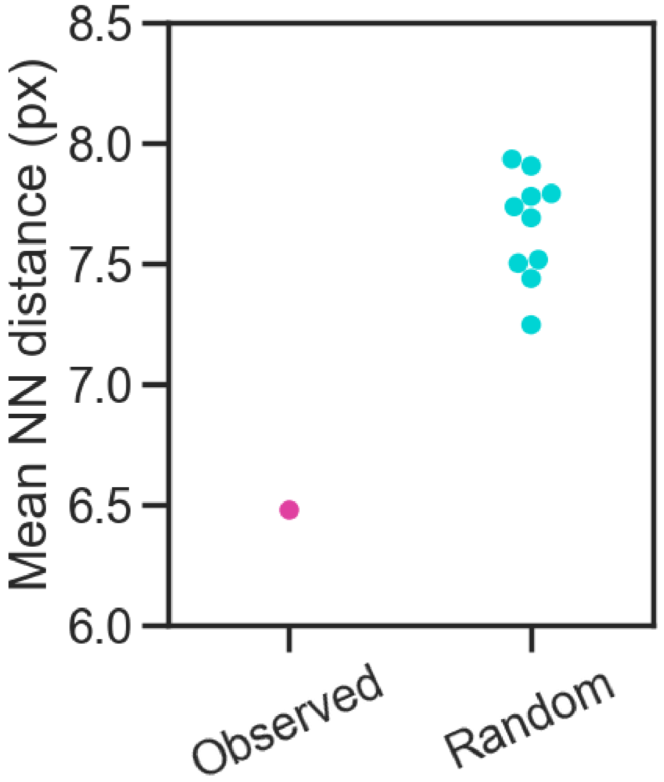
```
  y    x  nn_dist_px
 78  142    7.071068
103  159    5.656854
 94  142    8.544004
 94  157    2.236068
 79  151    3.162278
 87  151    2.236068
```



Clustering – are objects aggregated?



one run



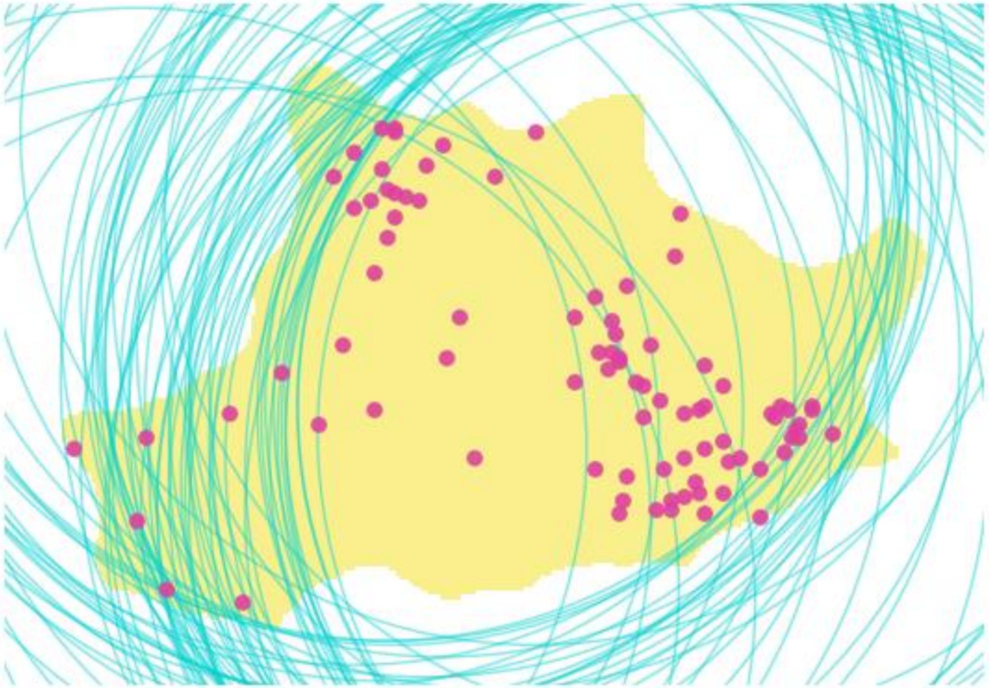
Clark-Evans index (CE):

$$CE = \frac{mean\ NN\ dist\ (observed)}{mean\ NN\ dist\ (random)} = \begin{cases} < 1: clustering \\ = 1:\ random. \\ > 1: dispersion \end{cases}$$

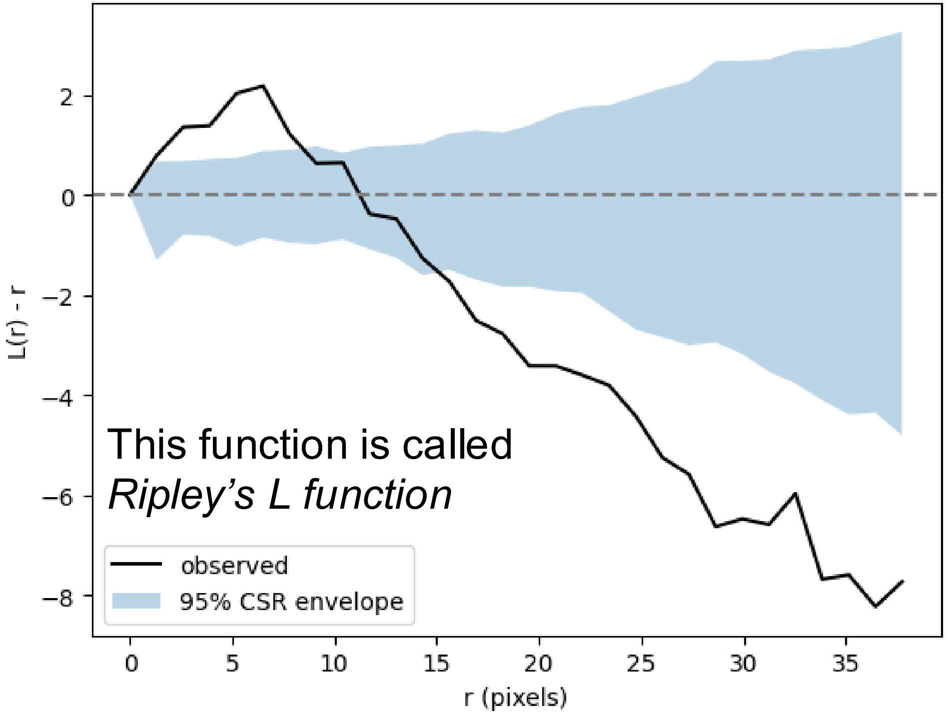
$$CE = \frac{6.5\ px}{7.8\ px} = 0.83$$

At which scale are objects aggregated?

For each spot count how many other spots are within radii of increasing size.



Plot Ripley's L against radius



This function is called *Ripley's L function*

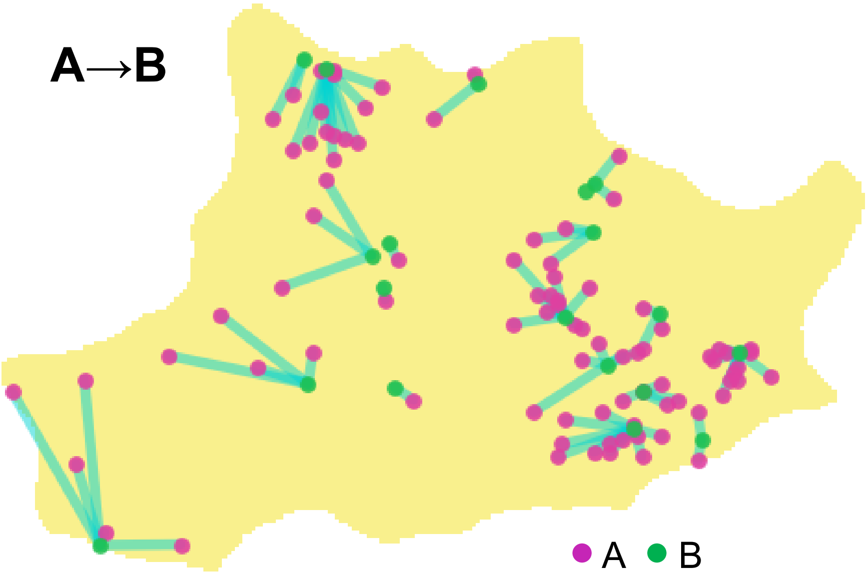
Calculate Ripley's L per radius

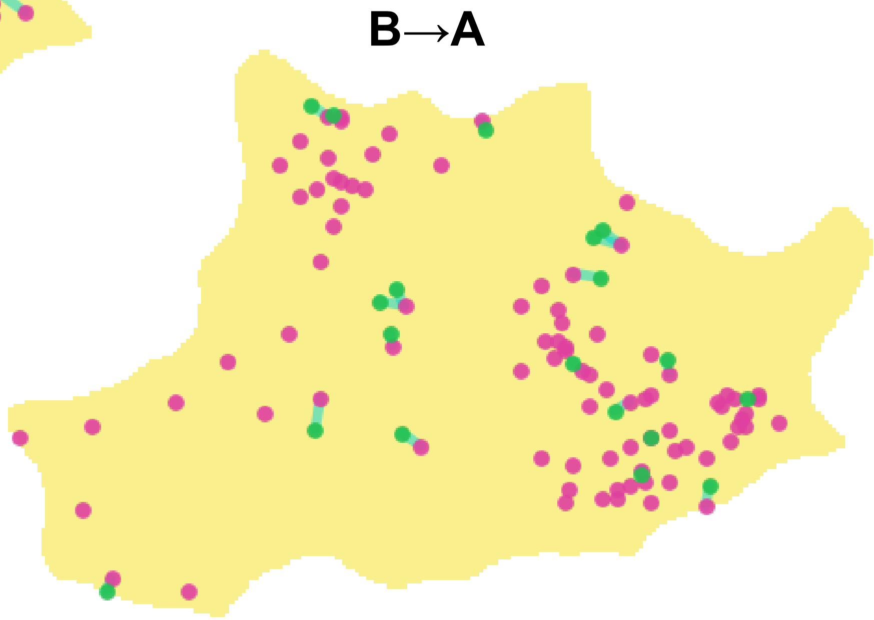
$$L(r) - r \propto \underbrace{N_{spots}(expected)}_{\text{actual count}} - \underbrace{N_{spots}(expected)}_{\text{density × area}} = \begin{cases} > 0: clustering \\ = 0: \ random. \\ < 0: dispersion \end{cases}$$

Interpretation

- observed curve within CSR envelope: not different from random
- observed curve outside CSR envelope: clustering/dispersion at that length scale

Cross NN distance measures distance between nearest neighbour across two classes of objects





Note:
Directionality matters!
Measuring A→B ≠ B→A

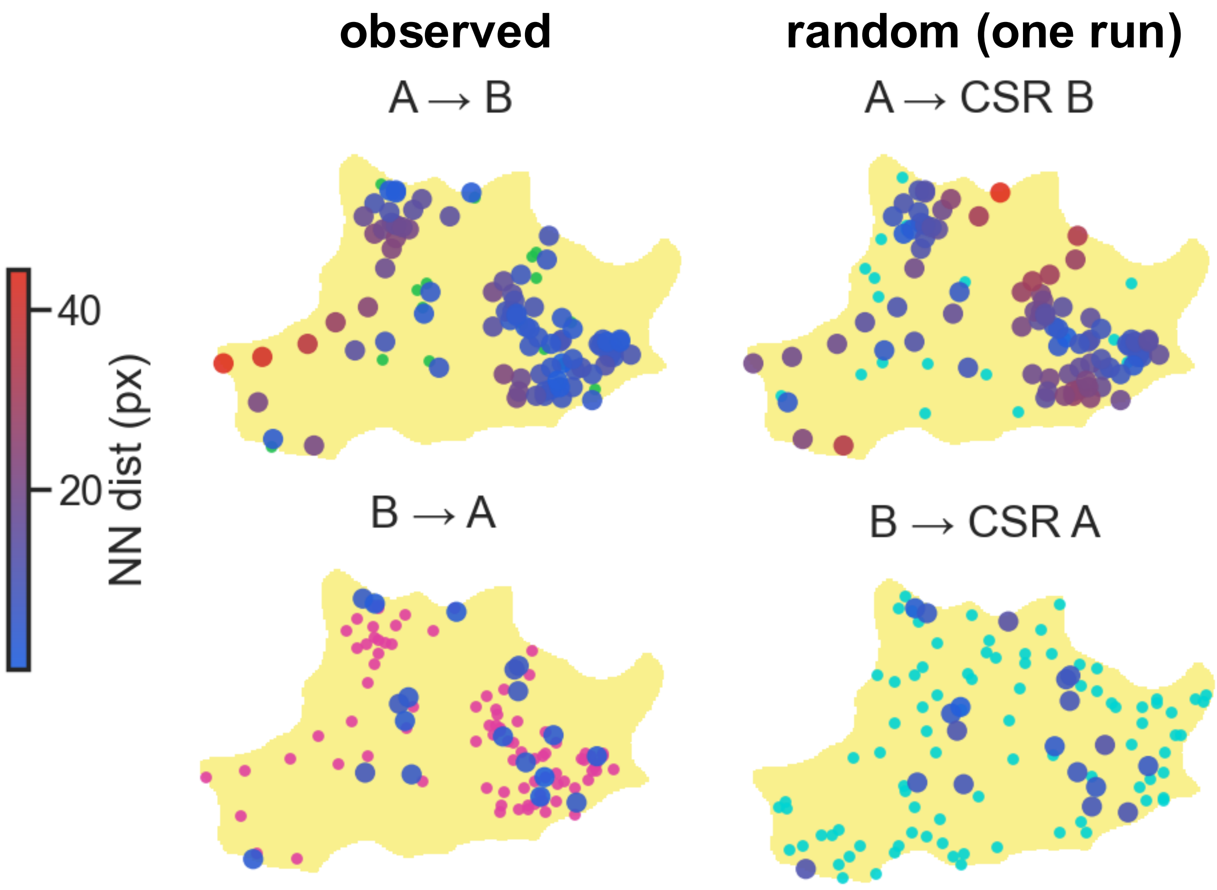
Algorithm
- For each spot
- find the nearest spot in the other distribution

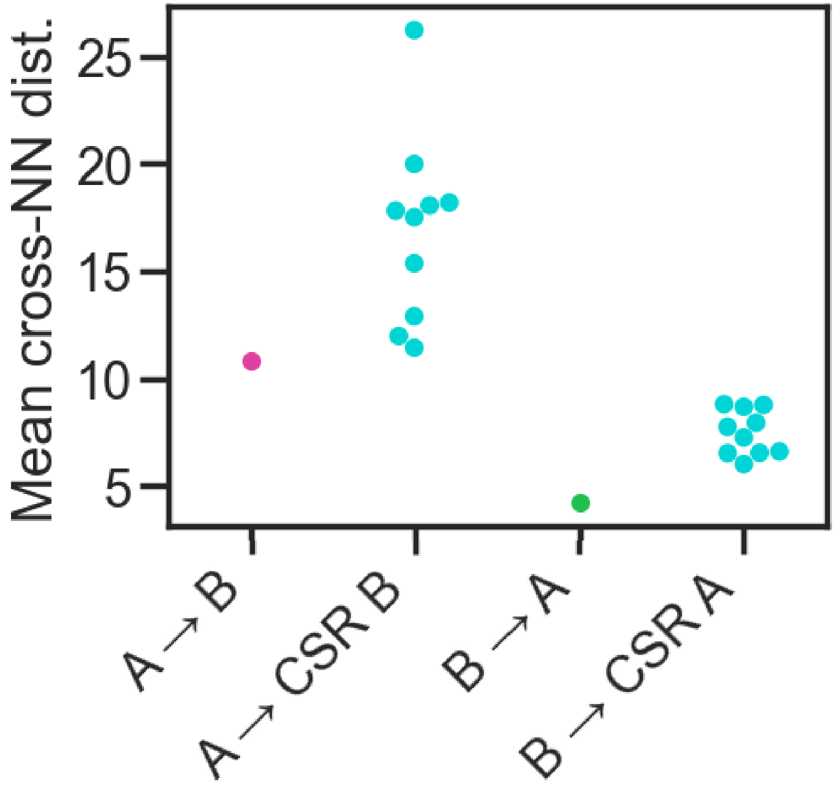
Variations
- Find the k nearest neighbors
- Radius queries: connect all spots within a specified radius

Benefits
- intuitive
- captures local density

Limitations
- Can under-connect points and produce sparse graphs
- Cross distance can be abundance dependent

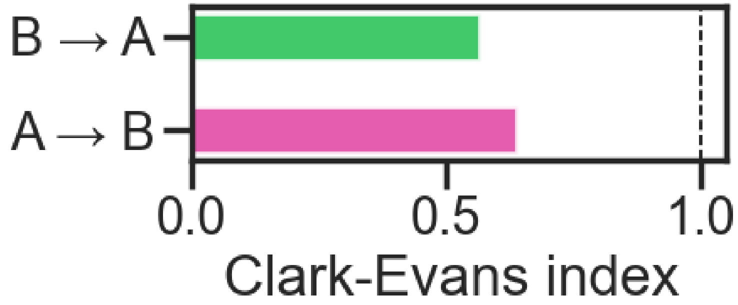
Cross-aggregation – are objects of two classes closer than random





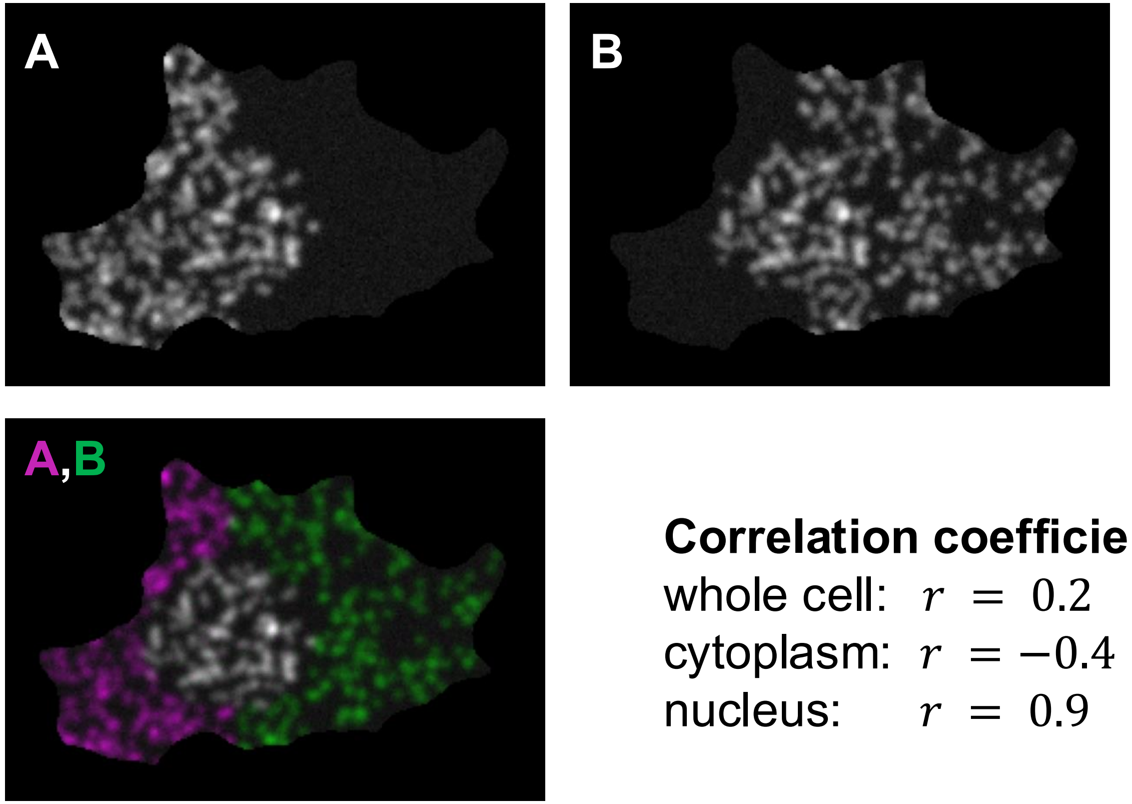
Clark-Evans index:

$$CE = \frac{mean\ NN\ dist\ (observed)}{mean\ NN\ dist\ (random)} = \begin{cases} < 1: clustering \\ = 1:\ random. \\ > 1: dispersion \end{cases}$$



Pixel-based co-localisation: Pearson’s correlation

Are Proteins A and B colocalized and if so, where?



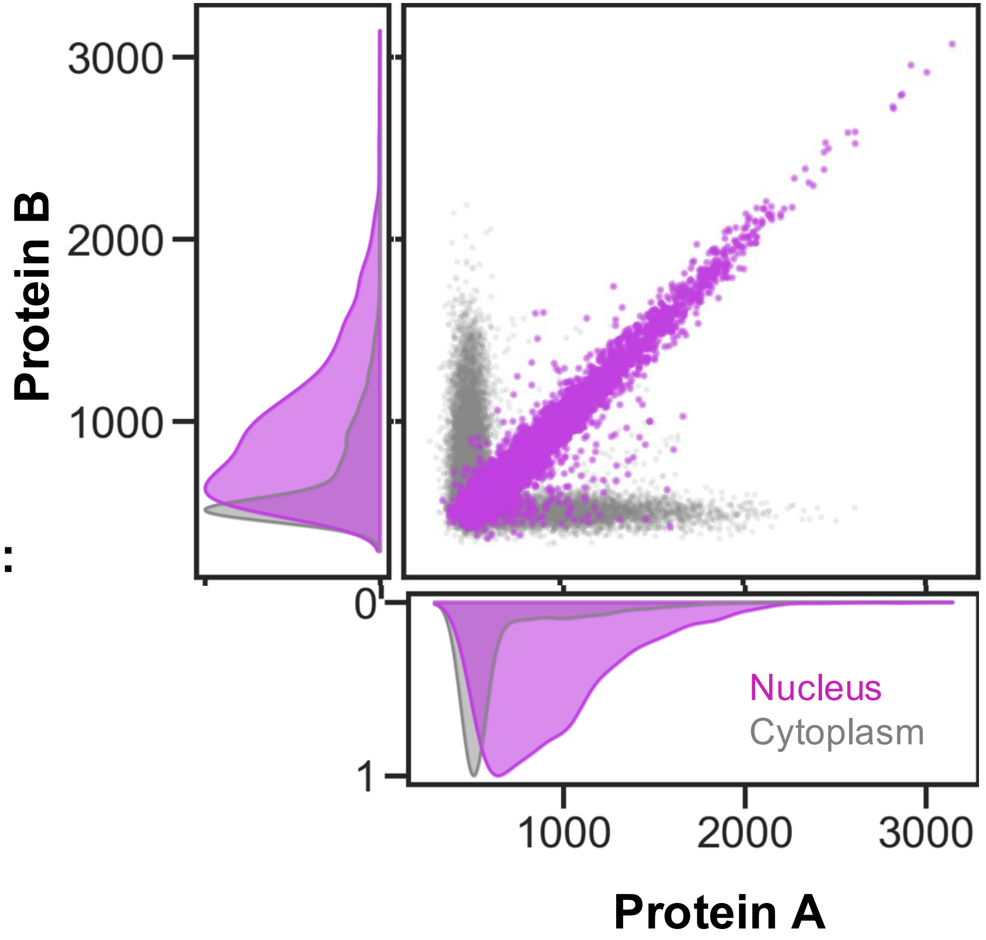
Correlation coefficients:

whole cell: $r = 0.2$

cytoplasm: $r = -0.4$

nucleus: $r = 0.9$

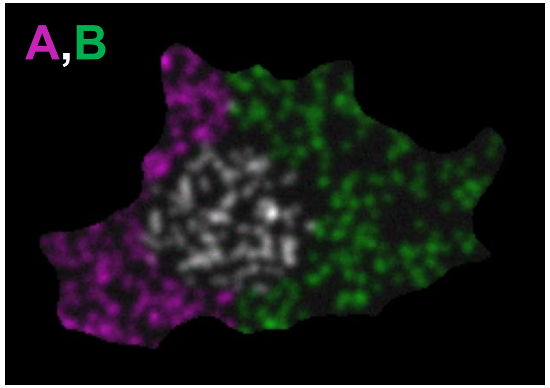
Pixel-based correlation between channels



“Pearson’s correlation reports on whether *two signals co-vary across a region*.”

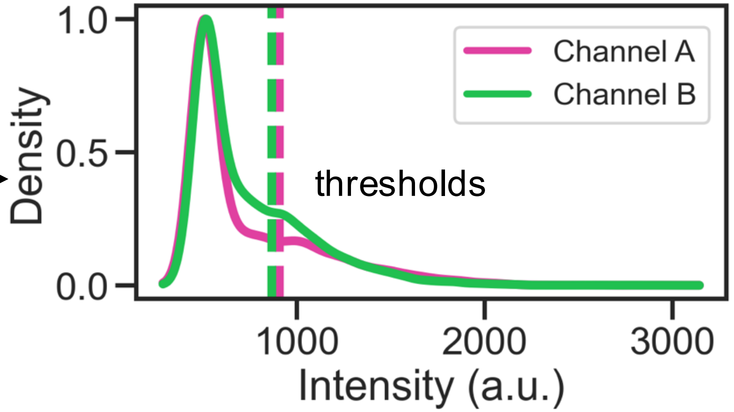
Area-based co-localisation:
Mander’s colocalization coefficients

What fraction of A and B are overlapping?



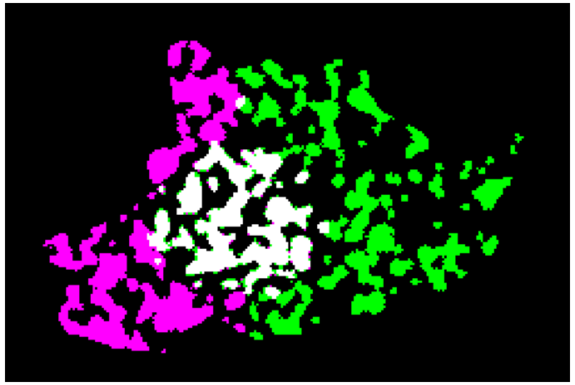
Threshold histogram

define positive/negative regions

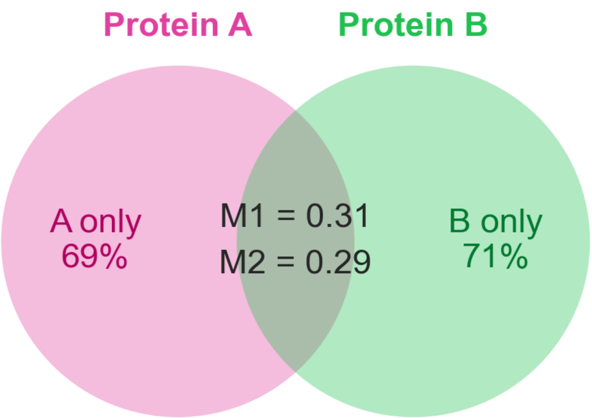


Two binary masks

with partial overlay

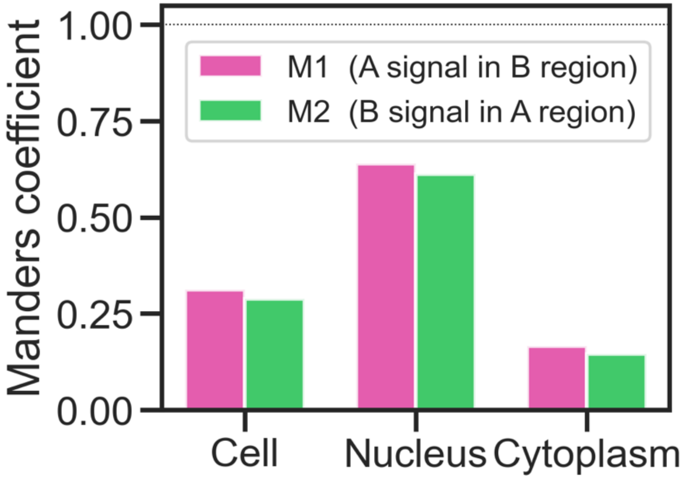


Quantify overlap per pixel



M1, M2 are Mander’s coefficients

Repeat for different regions



“Mander’s coefficients report, for a specified threshold, *how many pixels are positive for both signals as a fraction of either signal*.”

What determines what type of analysis you should do?

YOUR SCIENTIFIC QUESTION

What spatial analysis is good for

Spatial analysis answers questions that require knowing where things are!

How crowded, and does crowding matter?

Who is near whom?

Is something organized or random?

Do two signals/ objects coincide?

Where things are shapes what they can do!

What spatial analysis is not so good for

Demonstrating molecular interactions

“We can show that proteins or cells are close, but not that they interact.”

Deducing processes from patterns

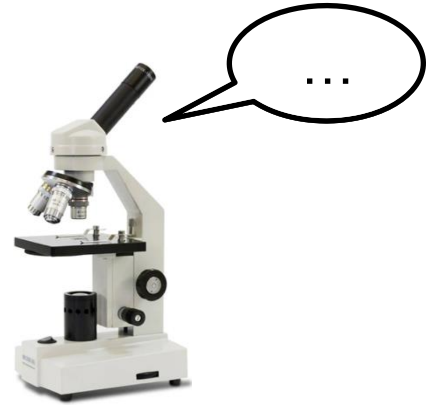
“We can show aggregation, but we can’t directly show the cause.”

This can only be solved by your experimental design.
(KOs, drugs, clever setups,…)

Scale-free answers

“We can show that proteins aggregate, but we need to specify the scale.”

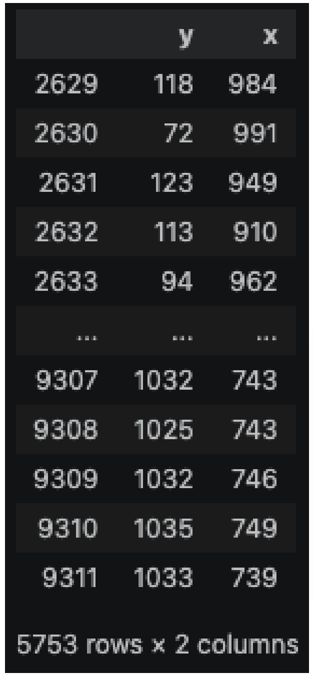
This can never be solved.



A typical spatial analysis pipeline

QUESTION:
Do my two favorite proteins colocalize?

Coordinates / Spots



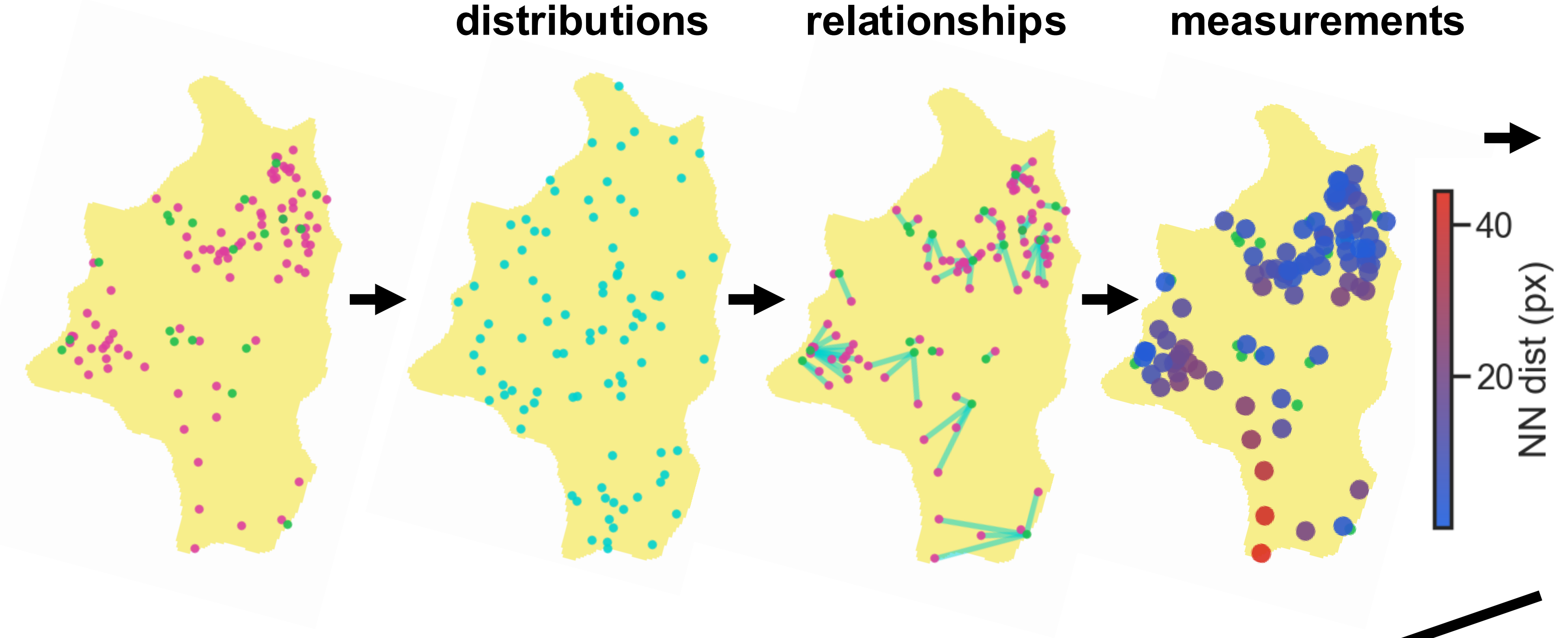
	y	x
2629	118	984
2630	72	991
2631	123	949
2632	113	910
2633	94	962
...	...	...
9307	1032	743
9308	1025	743
9309	1032	746
9310	1035	749
9311	1033	739

5753 rows × 2 columns

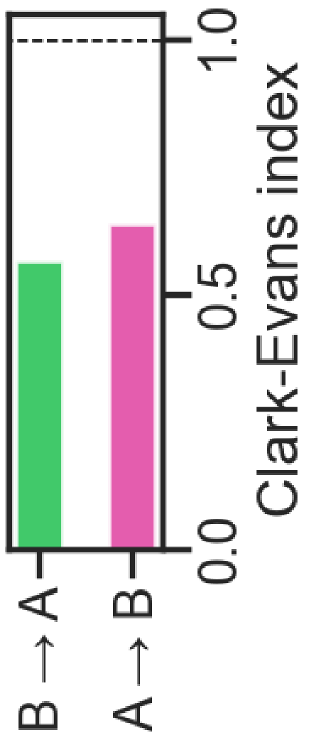
Define null distributions

Define spatial relationships

Define measurements



Results



CONCLUSION:
They localize closer to each other than expected by random chance.

Questions?