

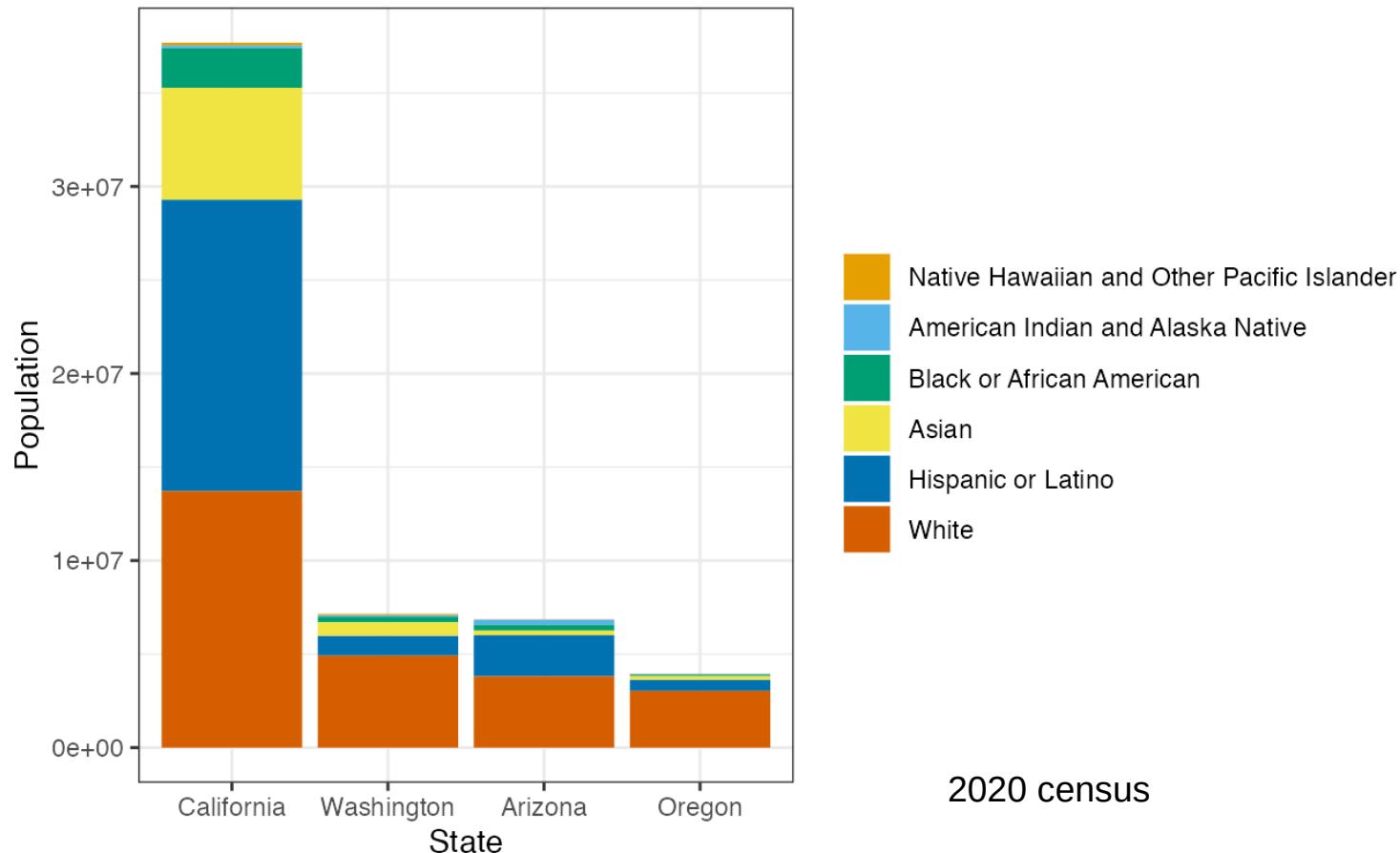
Exploratory spatial data analysis from single molecules to multiple samples

Lambda Moses
07/24/2024

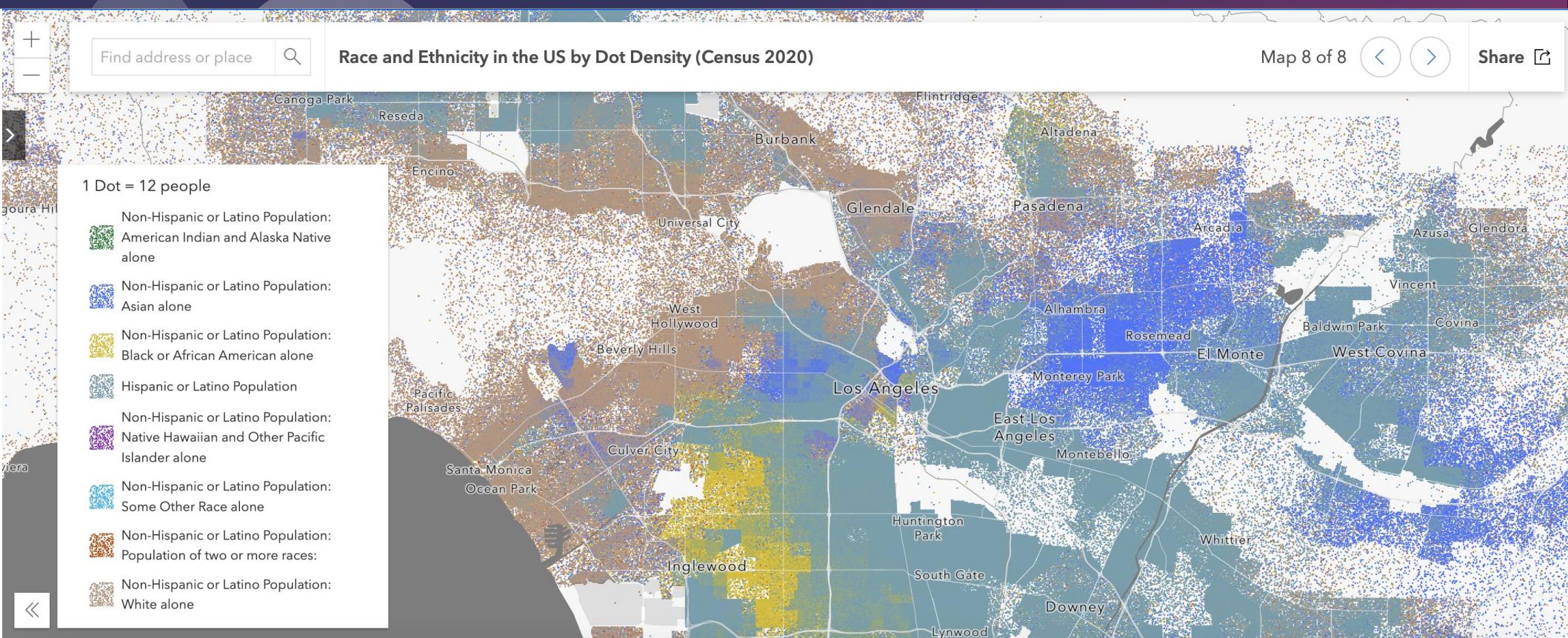
Outline

- Introduction: Philosophy of SpatialFeatureExperiment (SFE) and Voyager ESDA
- Getting, setting, and plotting fields of SFE objects
- Images and vector-raster interaction in SFE
- Spatial operations of SFE objects
- Spatial aggregation and ESDA across length scales
- Splitting and comparing between samples

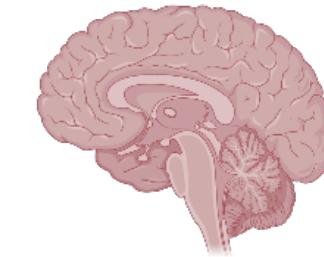
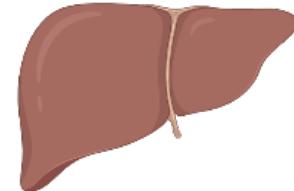
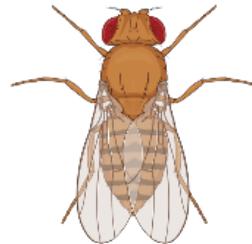
A geospatial analogy



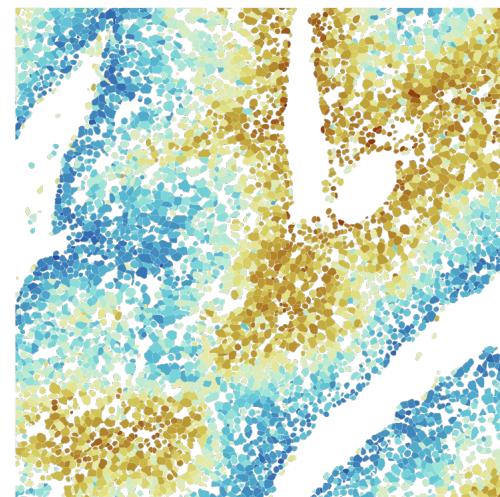
Many spatial questions to be asked



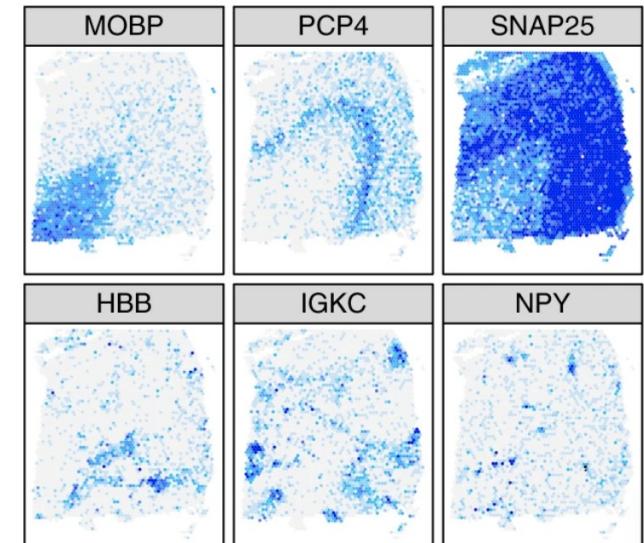
The whole is greater than the sum of the parts because Location! Location! Location!



BDGP

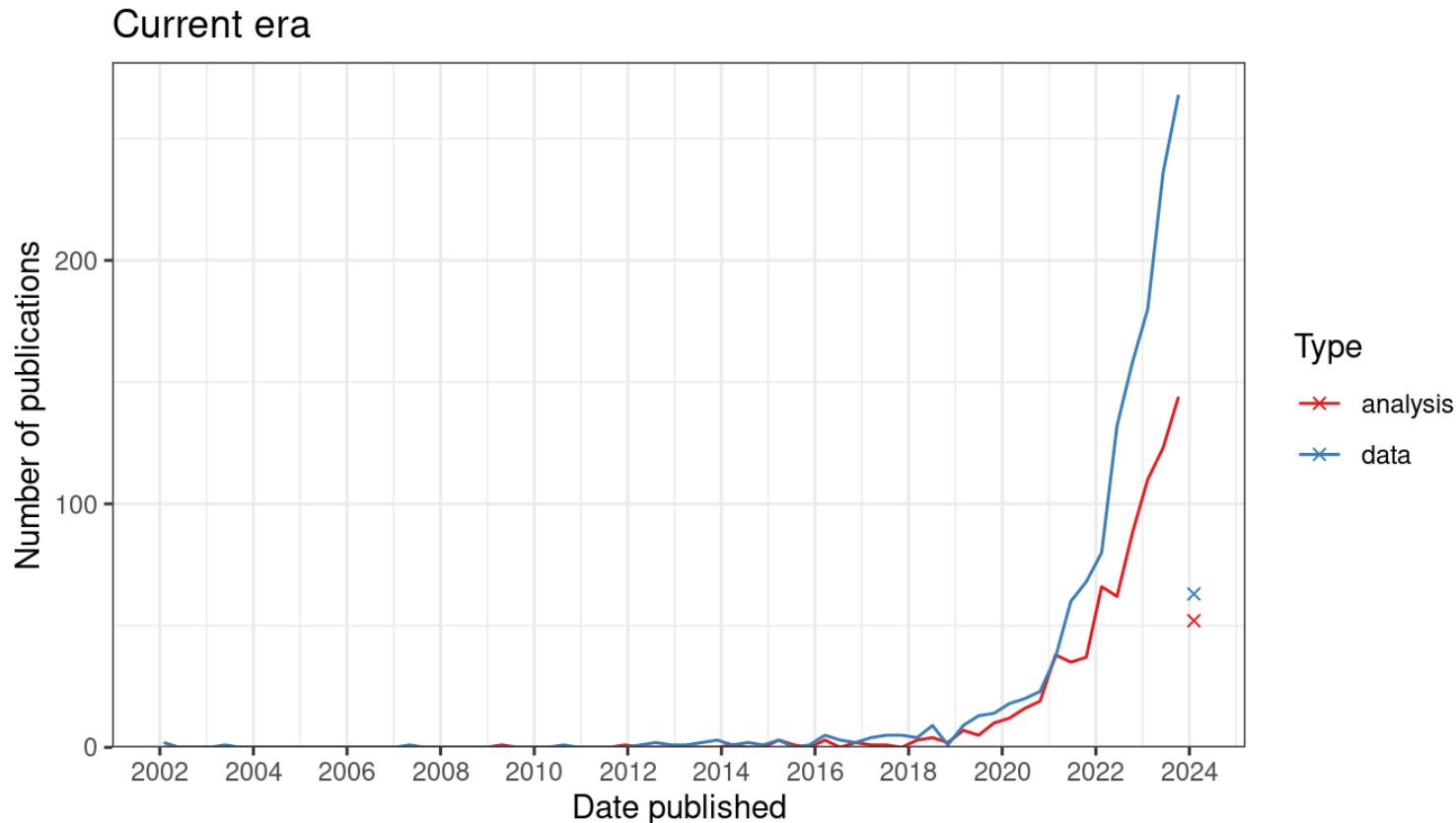


Voyager website

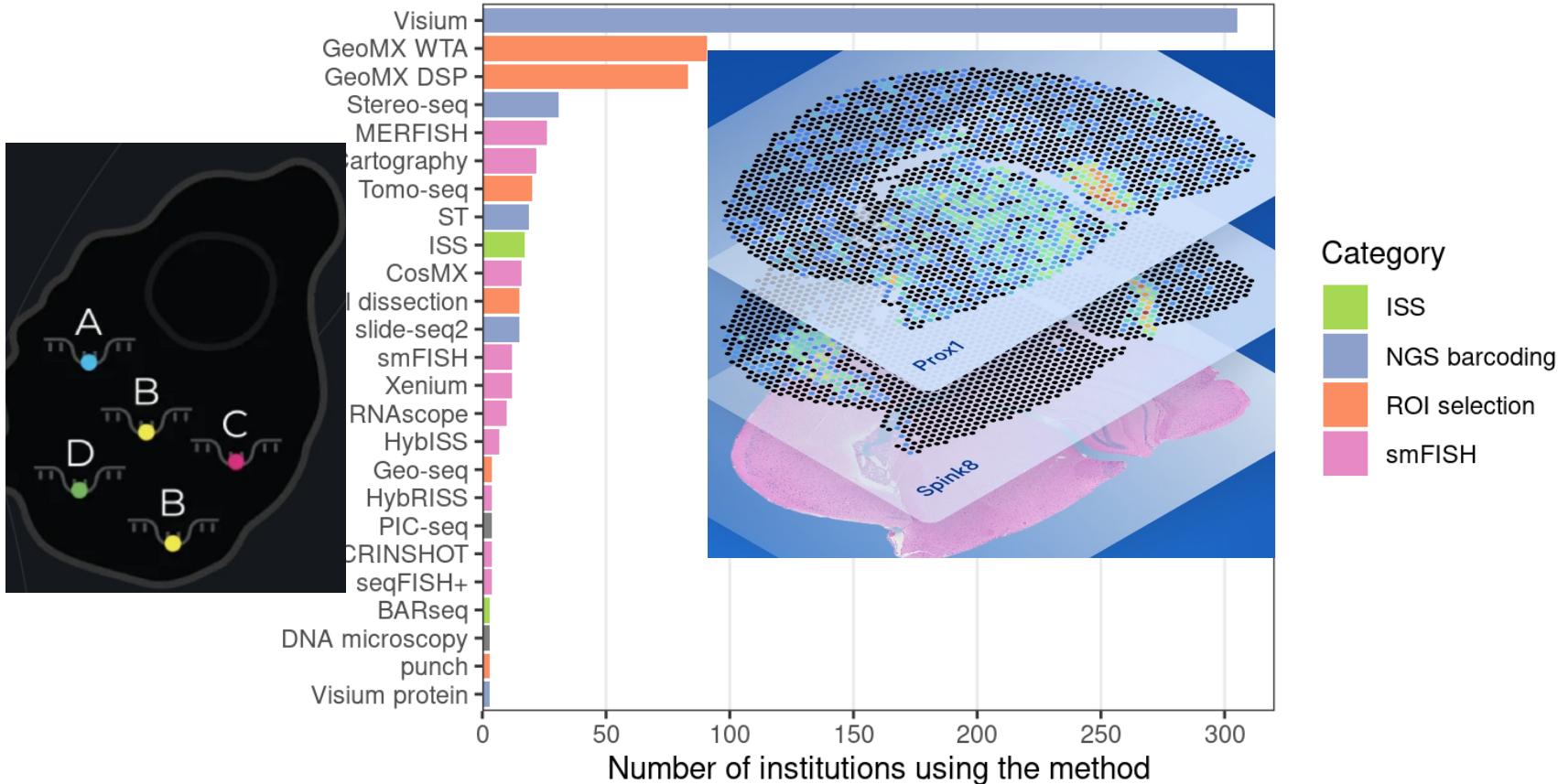


Maynard 2021 DLPFC

Drastic rise of spatial transcriptomics

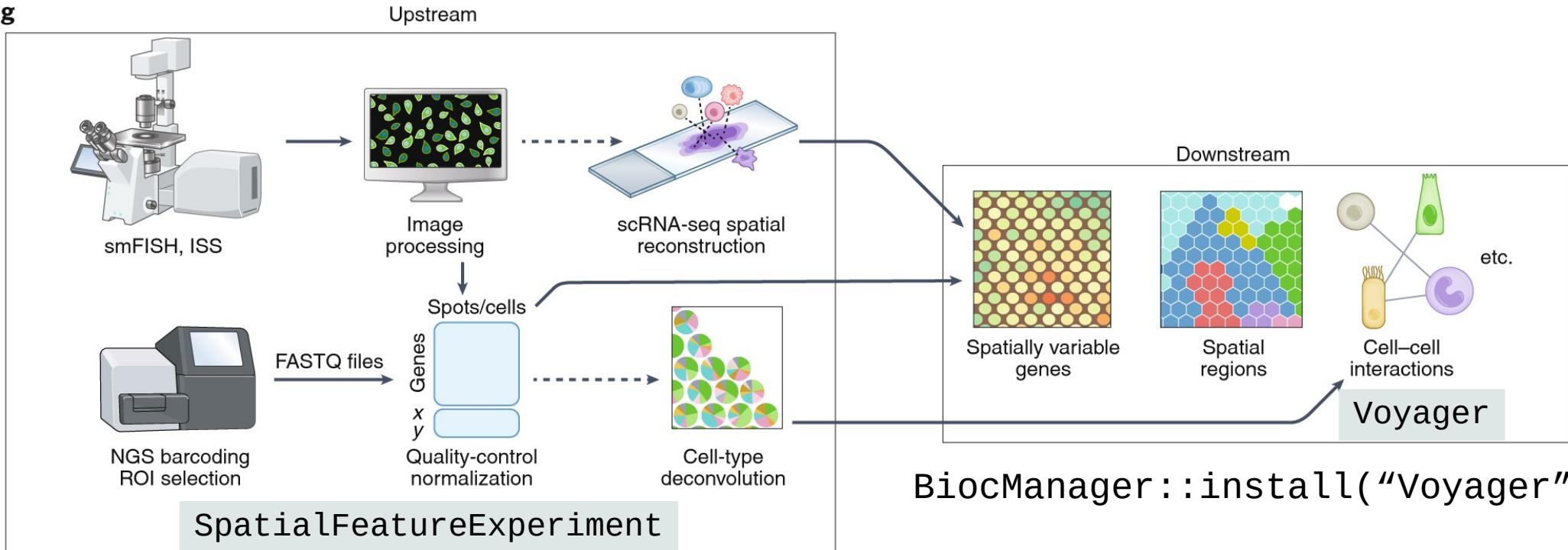


Drastic rise of spatial transcriptomics



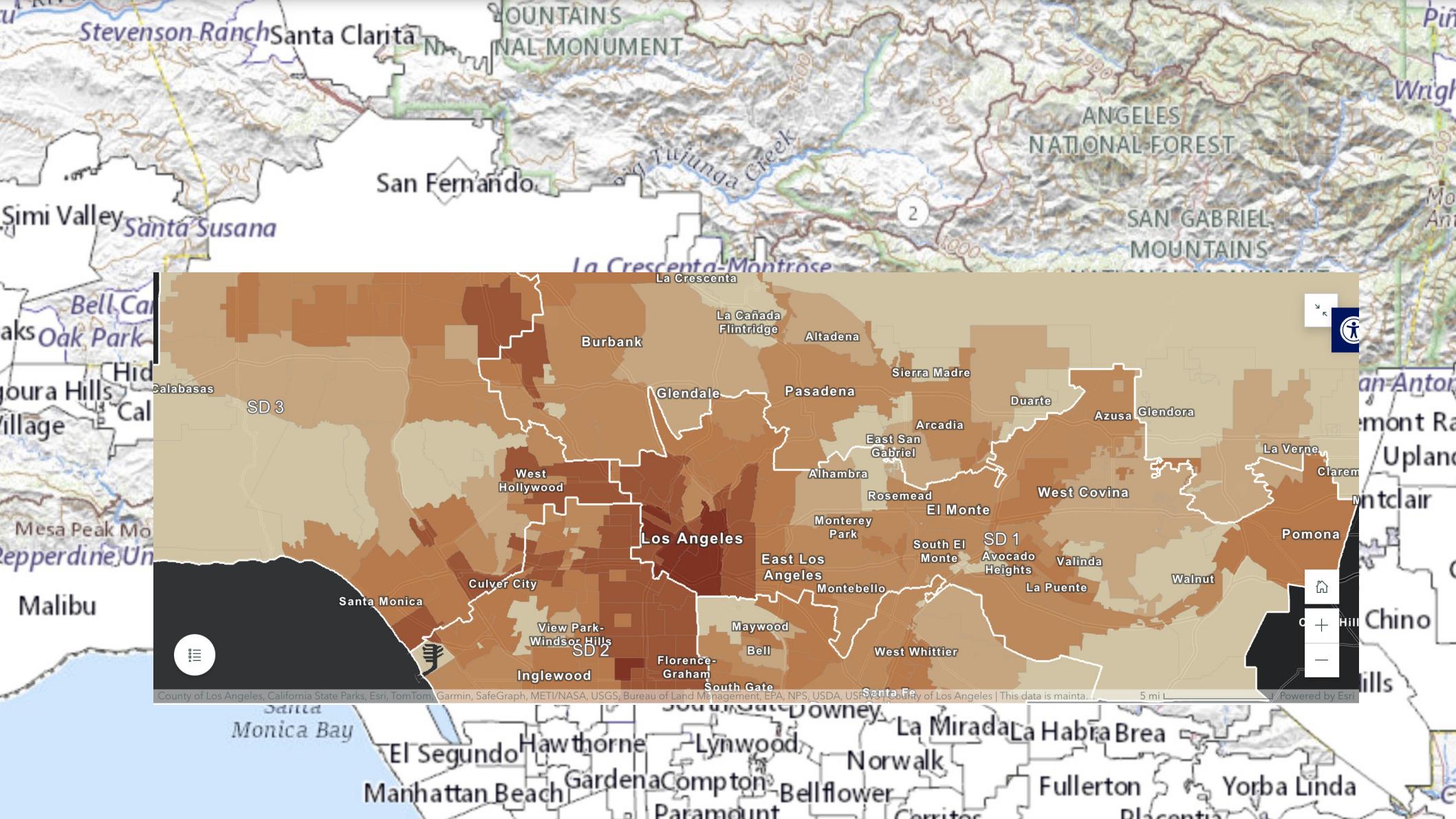
Spatial transcriptomics data analysis

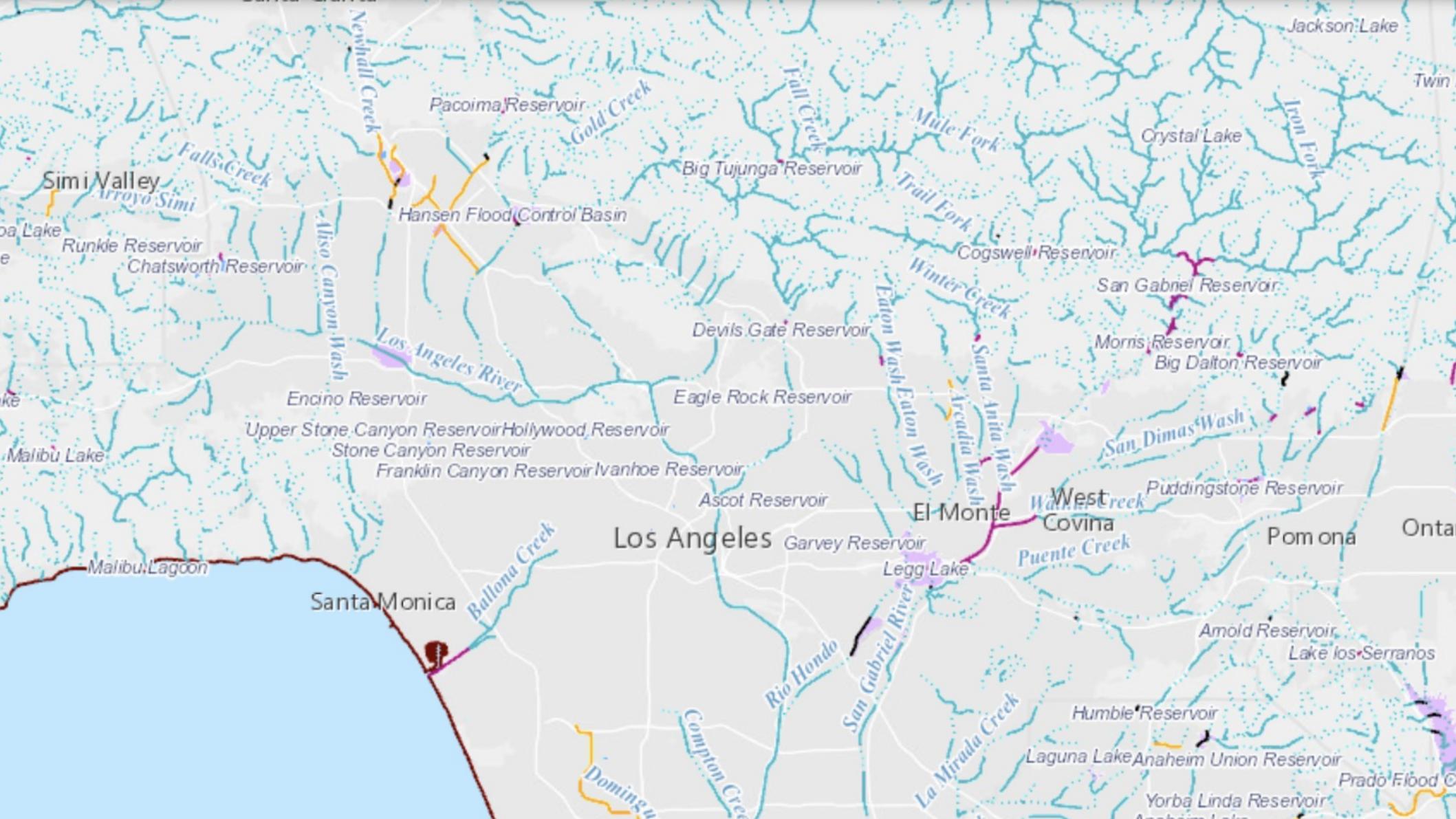
g

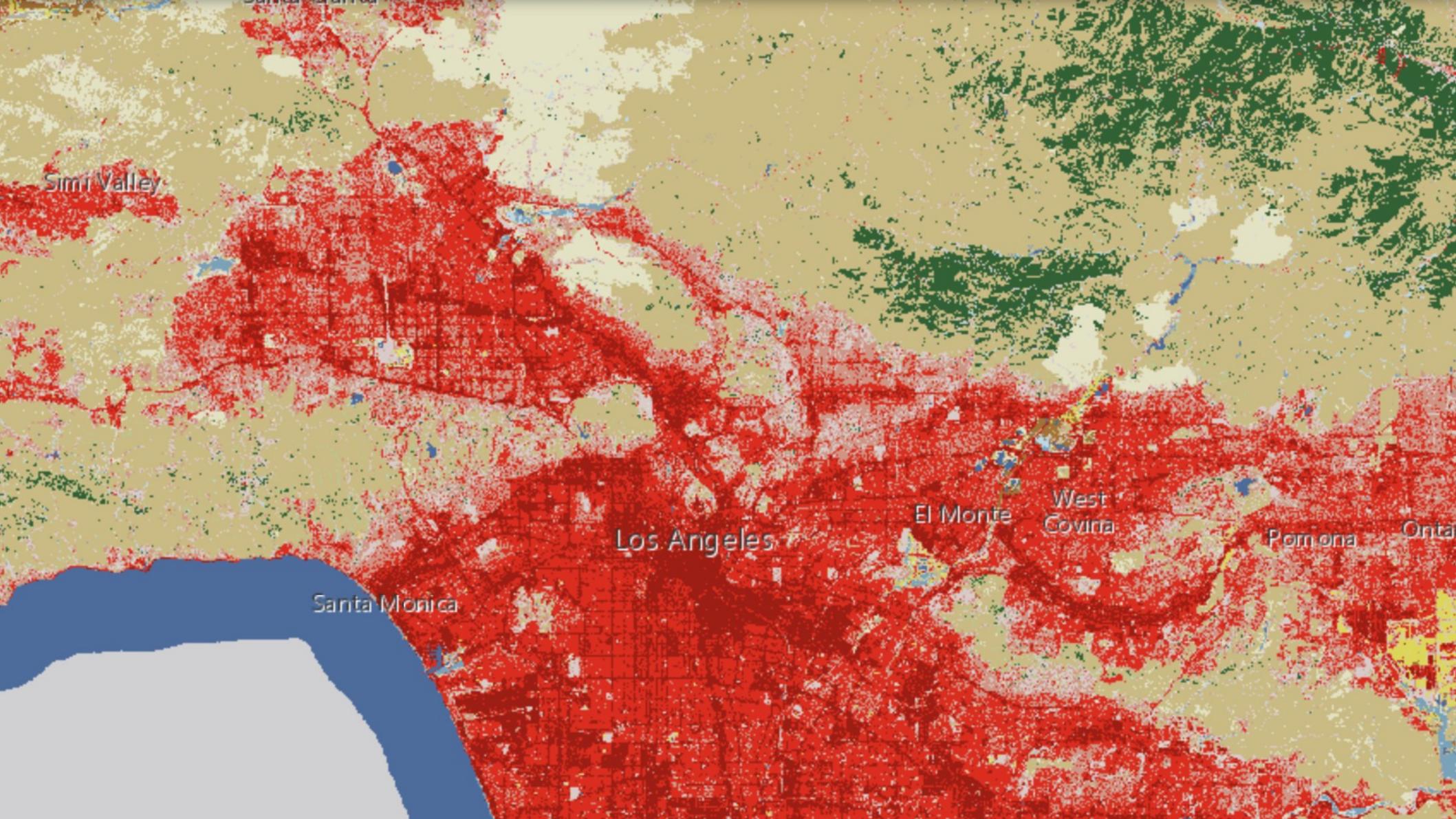


SpatialFeatureExperiment



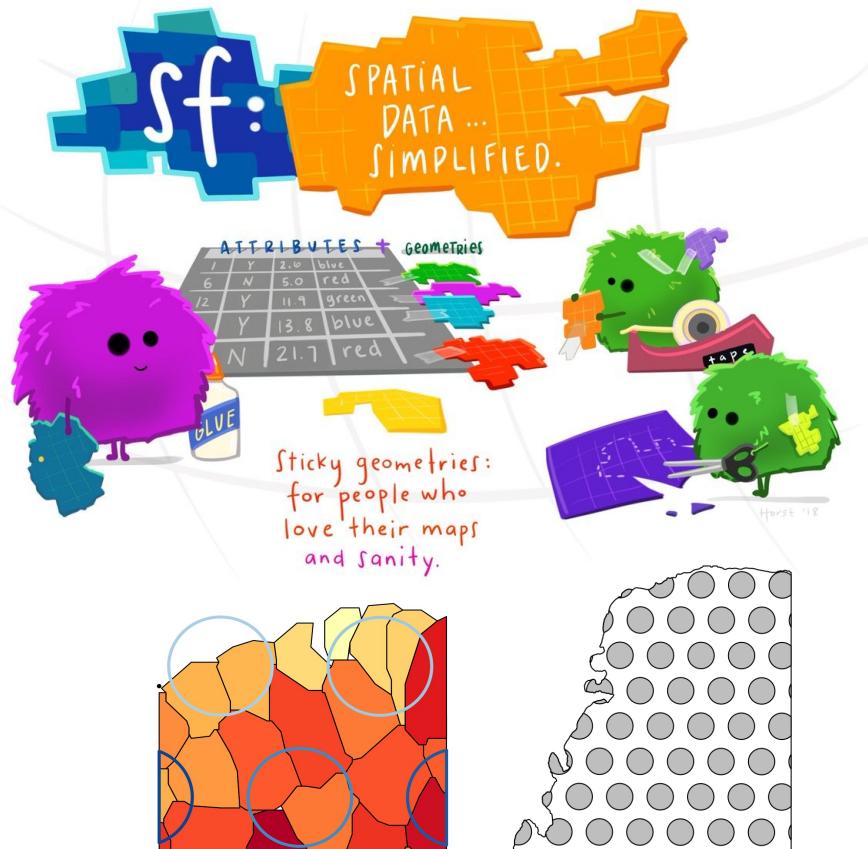








Simple Features: sf R package



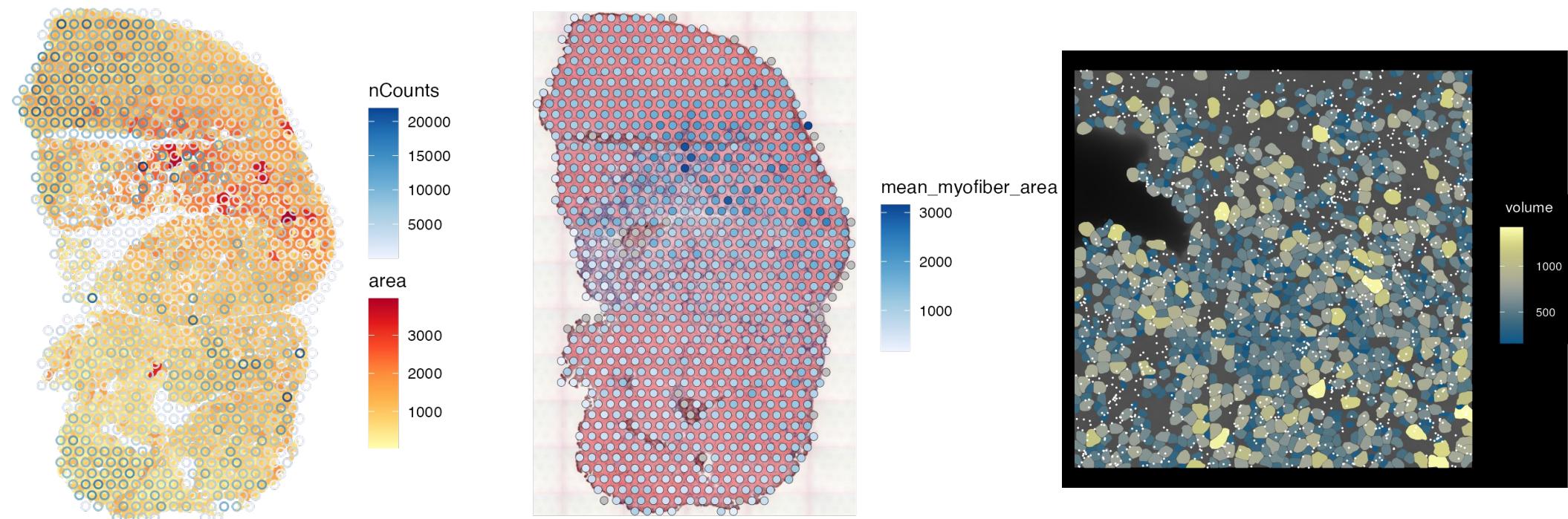
- Data frame with special column for vector geometries
- GEOS, GDAL C++ libraries for fast geometric operations
- Whether two geometries intersect
- Find the geometry of the intersections
- Find cells in each Visium spot
- Find cells or spots in each pathologist annotated region
- Crop SFE object by polygon
- Visualize cell morphology

Raster images: terra R package

- Data in pixels in arrays/images
- Can have any number of layers beyond RGB
- Spatially registered to coordinate system, e.g. longitude and latitude
- Use vector geometry to extract raster data
- Fast operations written in C++
- Out of memory operations



SpatialFeatureExperiment: layered maps in histological space

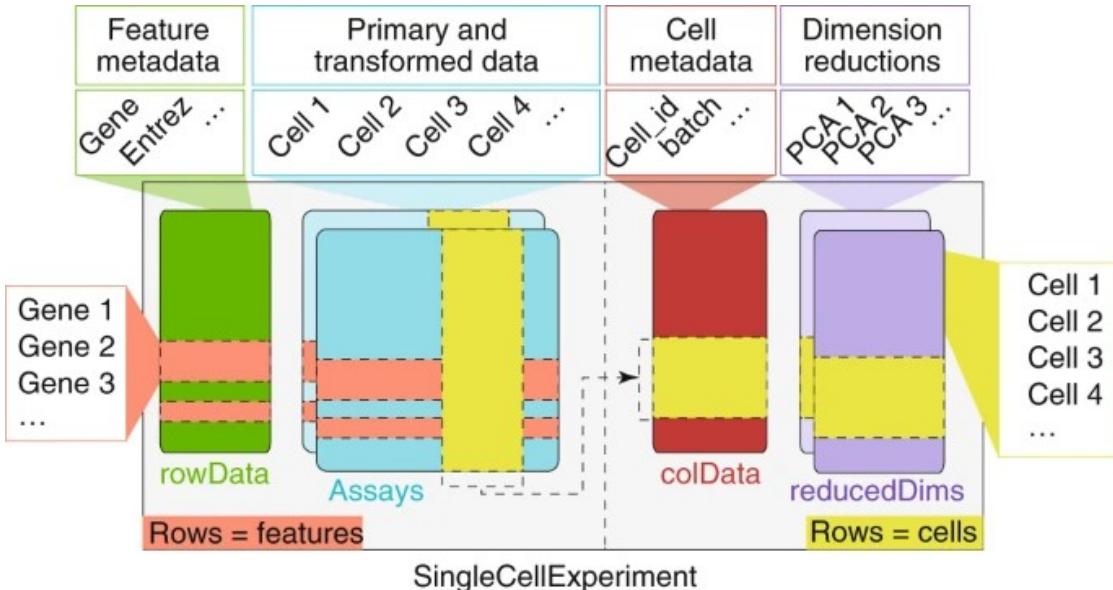


Data structure facilitates data analysis

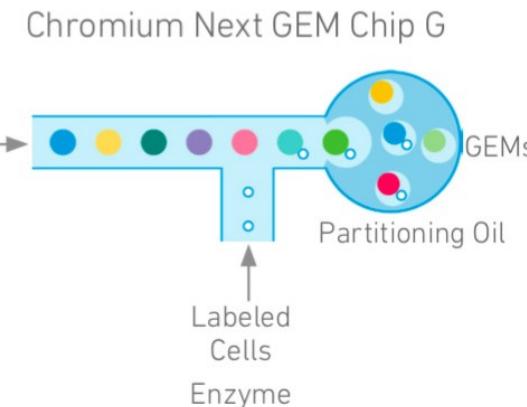
```
penguins
#> # A tibble: 344 × 8
#>   species island    bill_length_mm bill_depth_mm flipper_length_mm
#>   <fct>   <fct>           <dbl>            <dbl>                <int>
#> 1 Adelie  Torgersen      39.1            18.7                 181
#> 2 Adelie  Torgersen      39.5            17.4                 186
#> 3 Adelie  Torgersen      40.3            18                  195
#> 4 Adelie  Torgersen       NA              NA                  NA
#> 5 Adelie  Torgersen      36.7            19.3                 193
#> 6 Adelie  Torgersen      39.3            20.6                 190
#> # i 338 more rows
#> # i 3 more variables: body_mass_g <int>, sex <fct>, year <int>
```



SingleCellExperiment (SCE)



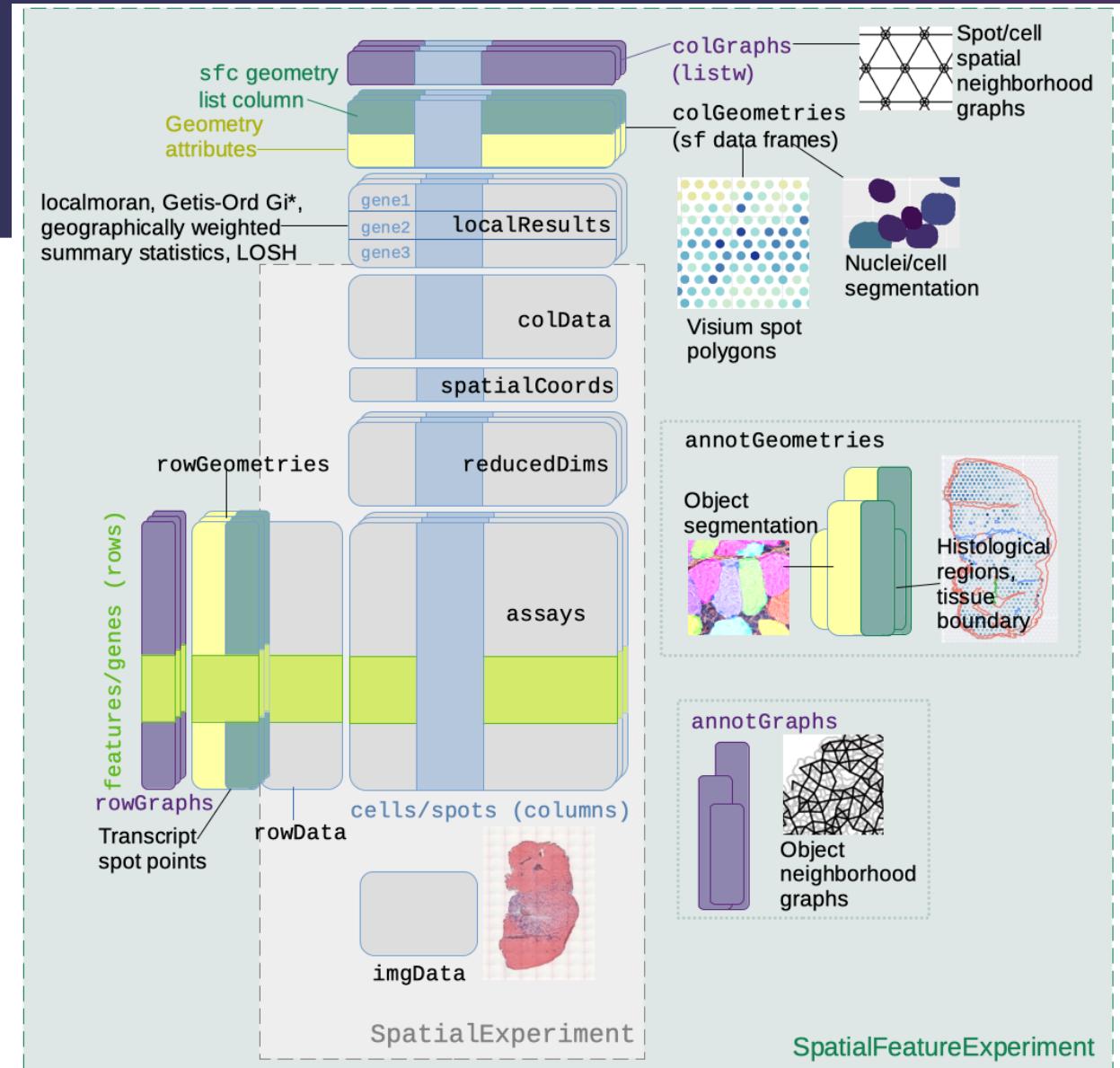
- Access existing scRNA-seq infrastructure for SCE
- SFE and Voyager are on Bioconductor:
 - Initial manual review
 - Extensive unit tests
 - Daily automated check



Amezquita R, Lun A, Becht E, Carey V, Carpp L, Geistlinger L, Marini F, Rue-Albrecht K, Risso D, Soneson C, Waldron L, Pages H, Smith M, Huber W, Morgan M, Gottardo R, Hicks S (2020). "Orchestrating single-cell analysis with Bioconductor." *Nature Methods*, 17, 137–145.
<https://www.nature.com/articles/s41592-019-0654-x>.

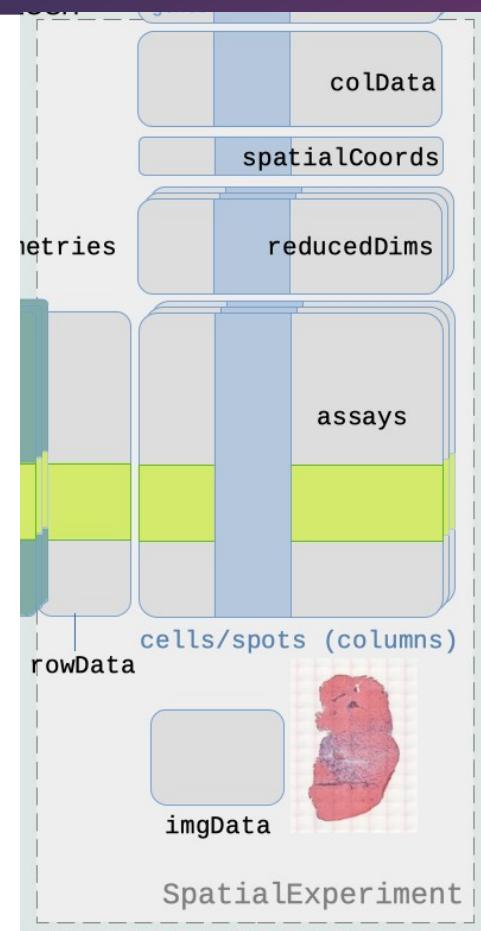
SpatialFeatureExperiment

- Rise in interest in spatial data structures: SeuratObject, newer AnnData, Staffli, Giotto, SpatialData
- SpatialExperiment (SPE): Existing package extending SCE for spatial - omics data
- SFE: data structure extending SPE with sf for vector, terra for raster
- Facilitate ESDA



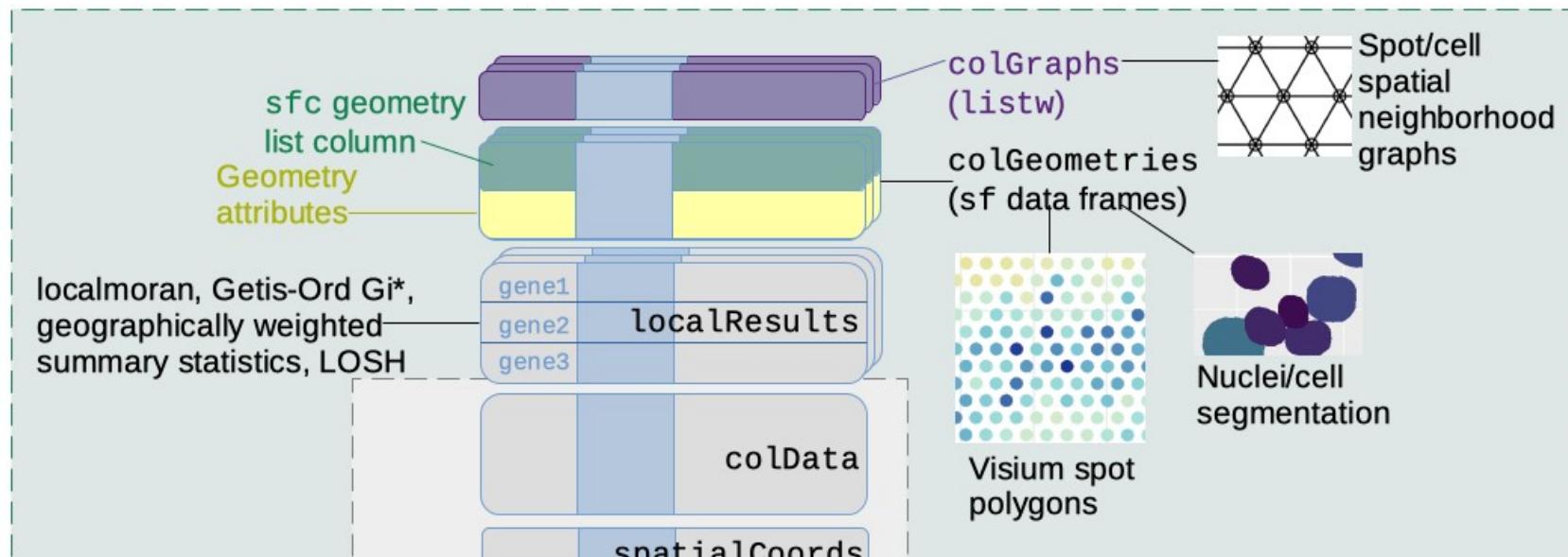
SpatialExperiment

- Existing Bioconductor extension of SCE for spatial -omics data
- **spatialCoords**: x, y coordinates of cell or spot centroids
- **imgData**: Histology image for data visualization



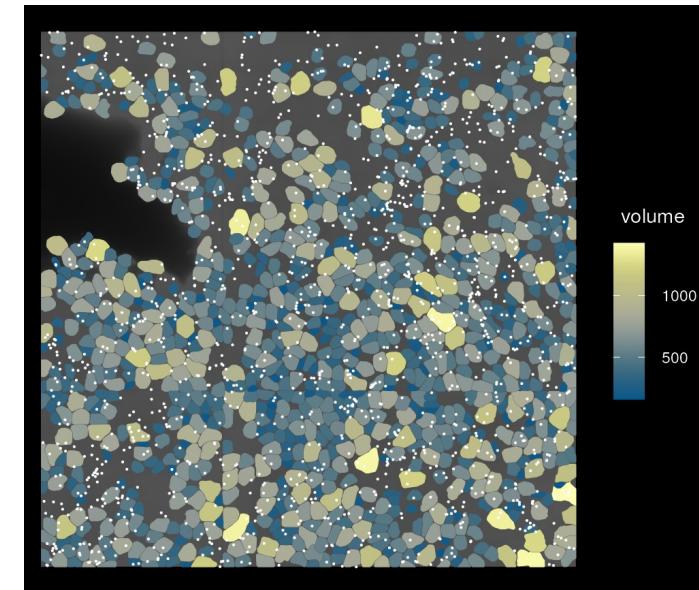
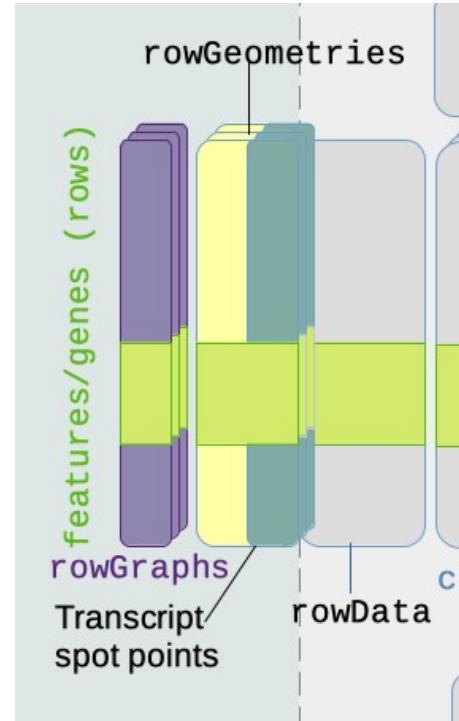
Column spatial information

- Corresponding to entities represented by the columns of the gene count matrix: Visium spots, cells and nuclei in MERFISH
- colGraphs: Indicate presence and strength of spatial dependence among these entities
- localResults: Store local spatial analysis results per gene/feature



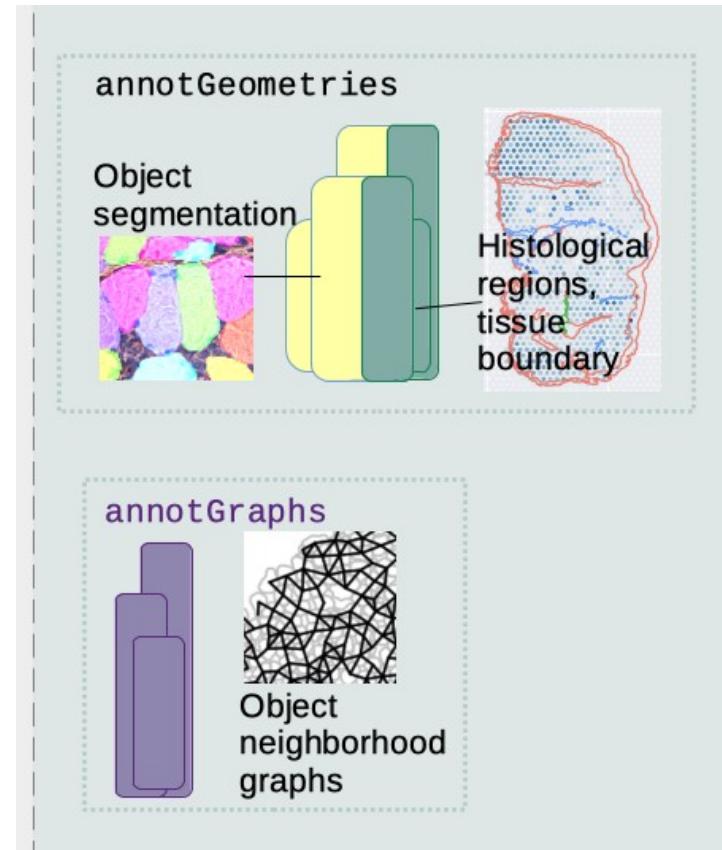
Row spatial information

- Corresponding to entities represented by rows of the gene count matrix: genes, features
- rowGeometries: Transcript spot points from smFISH
- rowGraphs: Potentially gene colocalization graph, but not currently used
- rowData: Gene symbols and annotation, feature QC, univariate global spatial analysis results



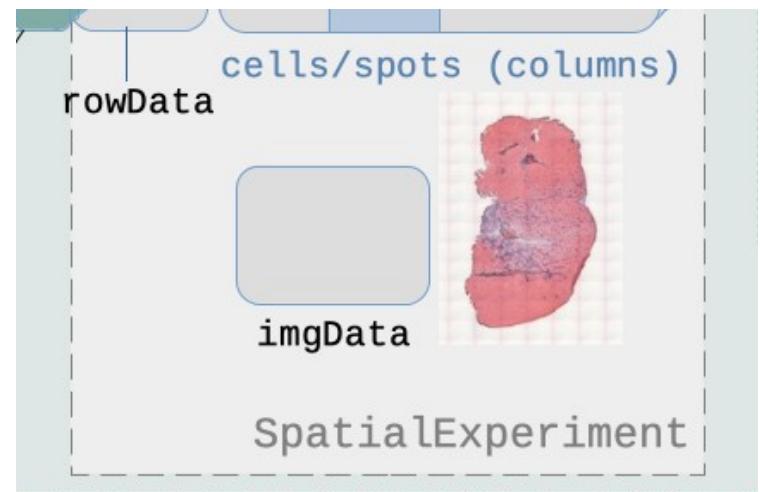
annotGeometries and annotGraphs

- Entities relevant to the dataset but don't correspond to columns of gene count matrix
- Myofibers, cells, and nuclei in Visium dataset
- Tissue boundary
- Pathologist annotated histological regions



imgData

- SpatRasterImage (terra, vector-raster interaction)
- BioFormatsImage (out of memory)
- ExtImage (image processing)
- Extent is required so the image is spatially registered with the rest of the data
- Extent: more general and flexible than SPE's scale factor

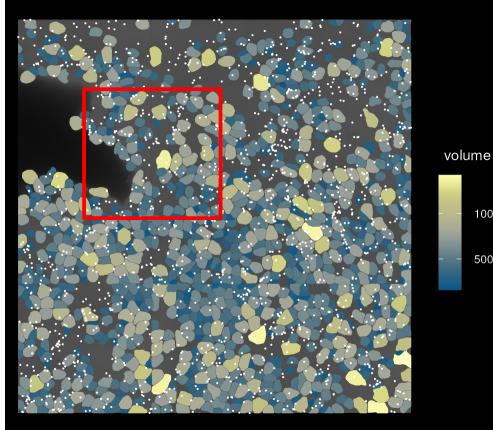


Operations on SFE objects

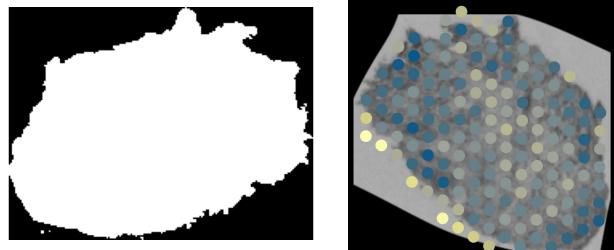
`sfe[i, j]`

`cbind(sfe1, sfe2)`

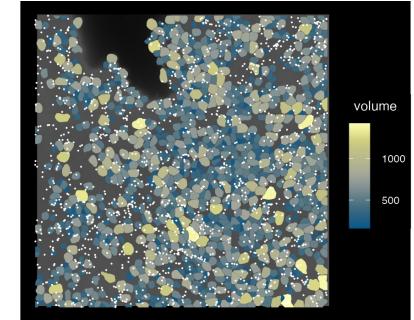
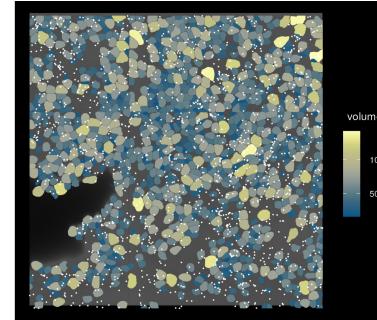
Cropping



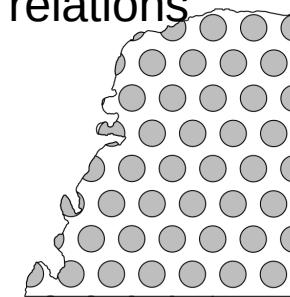
Vector-raster interaction



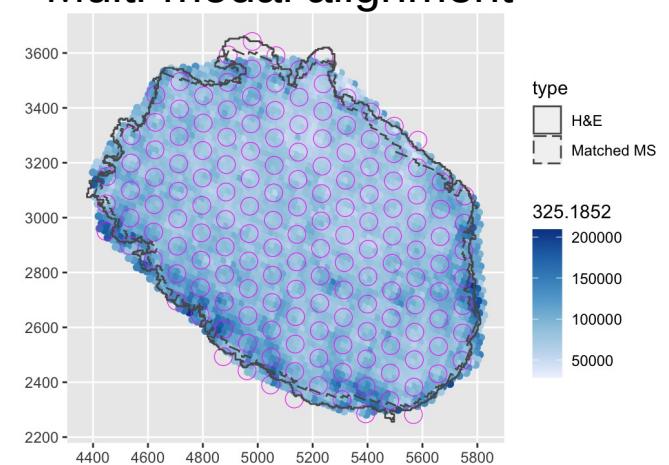
Affine transformations



Geometric relations



Multi-modal alignment



Future directions

- On disk geometry operations with sedona
- Findable, accessible, interoperable, and reusable (FAIR) principles
- MultiAssayExperiment extension of SFE for multi-omics data
- Multi-omics vignette

Preprint

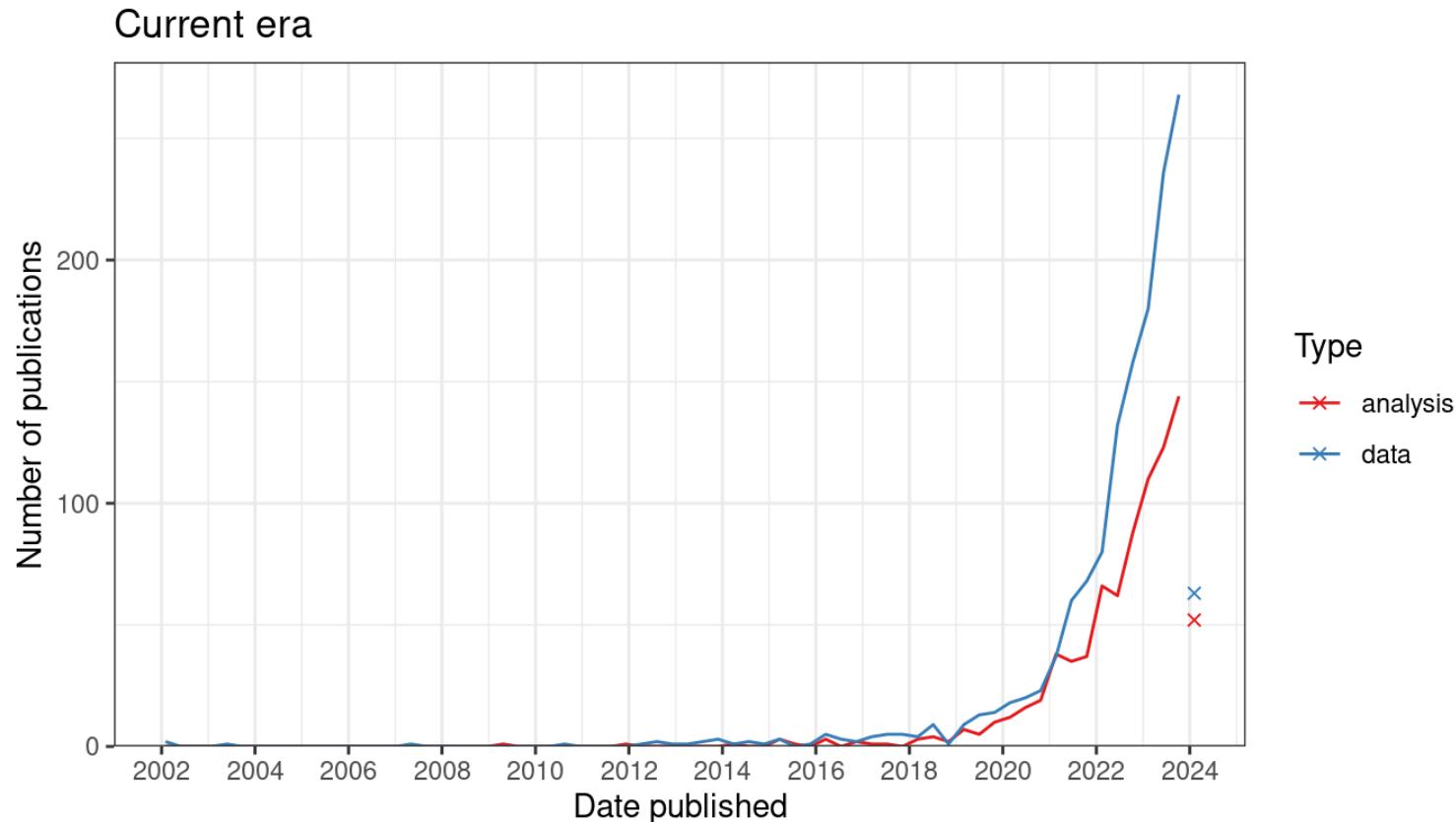


SCAN ME

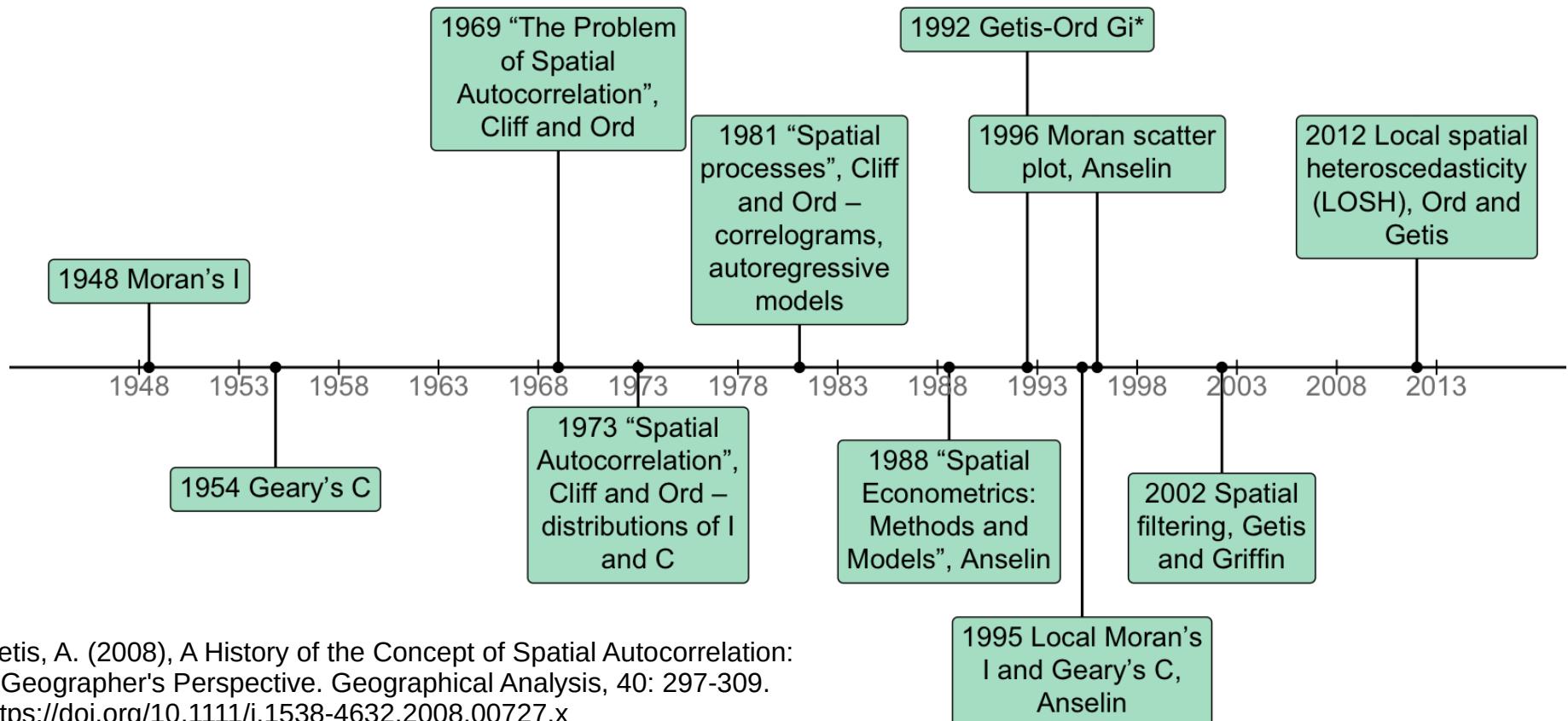
GEOGRAPHICAL FROM
SPATIAL TO
VOMAGER
SPATIAL-OMICS



Number of new publications in spatial transcriptomics over time



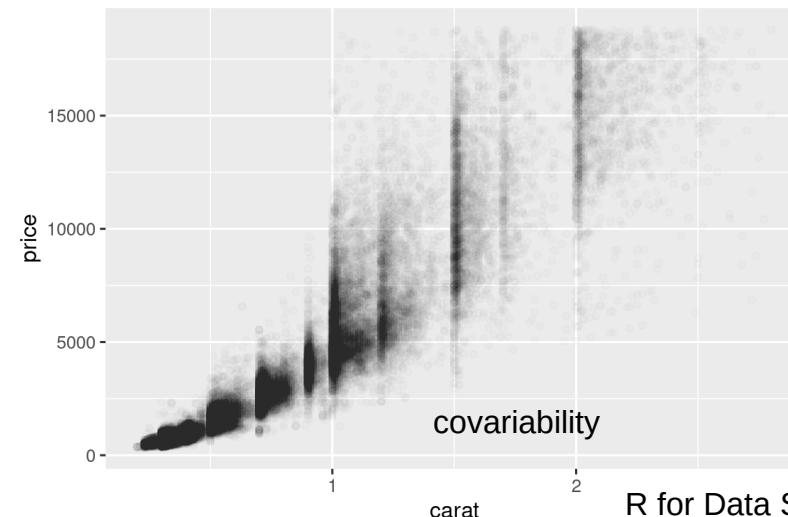
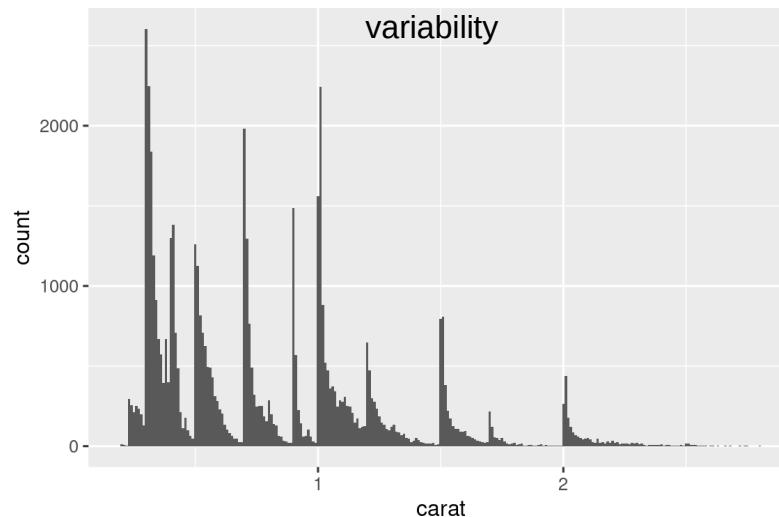
Timeline of spatial statistics in Voyager



Exploratory data analysis (EDA)

It is important to understand what you CAN DO before you learn to measure how WELL you have seem to DONE it (Exploratory Data Analysis, Tukey 1977)

“Far better an approximate answer to the right question, which is often vague, than an exact answer to the wrong question, which can always be made precise.” — John Tukey



R for Data Science, 2nd edition

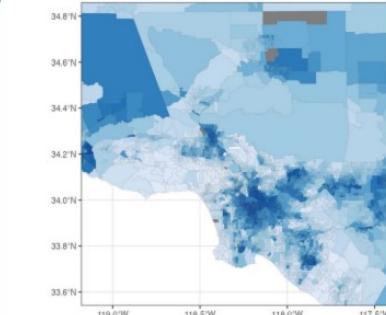
Exploratory spatial data analysis (ESDA): EDA for spatial aspects of the data, especially spatial autocorrelation

From geospatial to spatial transcriptomics

- Spatial point process: trendsceek, pciSeq, Bento, squidpy, spicyR
- Geostatistical (Gaussian process): SpatialDE, SPARK, GPcounts, BOOST-GP, spatialGE
- Moran's I: MERINGUE, Seurat, benchmarking SVG methods
- Local Moran's I: MERINGUE
- Lee's L: MERINGUE, Giotto
- Potts model spatial clustering: smfishhmrf (Giotto), BayesSpace
- GeoTIFF: samui browser
- Existing packages tend to focus on “how WELL you have seem to DONE it”
- Voyager: general ESDA (“what you CAN DO”), systematic uniform user interface, more efficient implementations

VOYAGER

From geospatial



spdep, PySAL:
spatial autocorrelation
gstat: geostatistical variograms



Raster image operations



Vector geometry operations



runUnivariate()
Moran's I,
Geary's C,
local Moran,
Getis-Ord Gi*,
correlogram,
variogram...

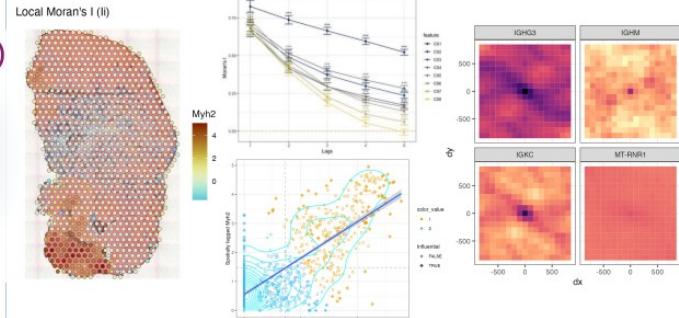
runBivariate()
Lee's L,
cross variogram

runMultivariate()
MULTISPATI PCA,
multivariate local C



to spatial -omics

Visualize data and spatial analysis results



Comprehensive and reproducible documentation

Tutorials with data from:

Visium

Slide-seq

Xenium

CosMX

Automated website build



SFEData package: example datasets



Compatibility tests

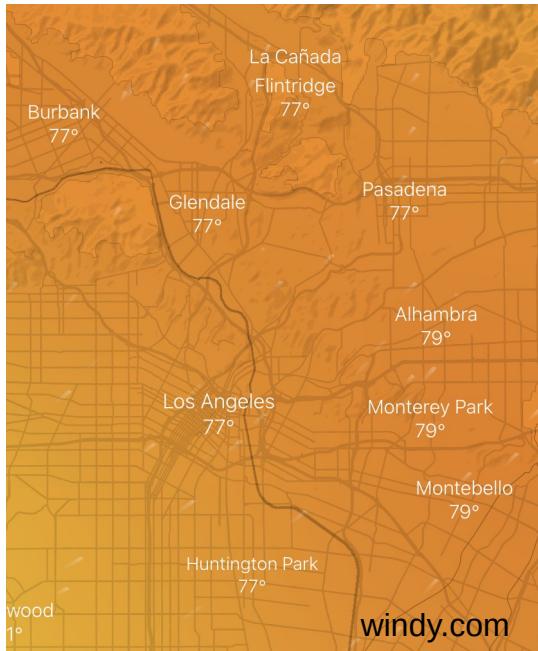


SCAN ME

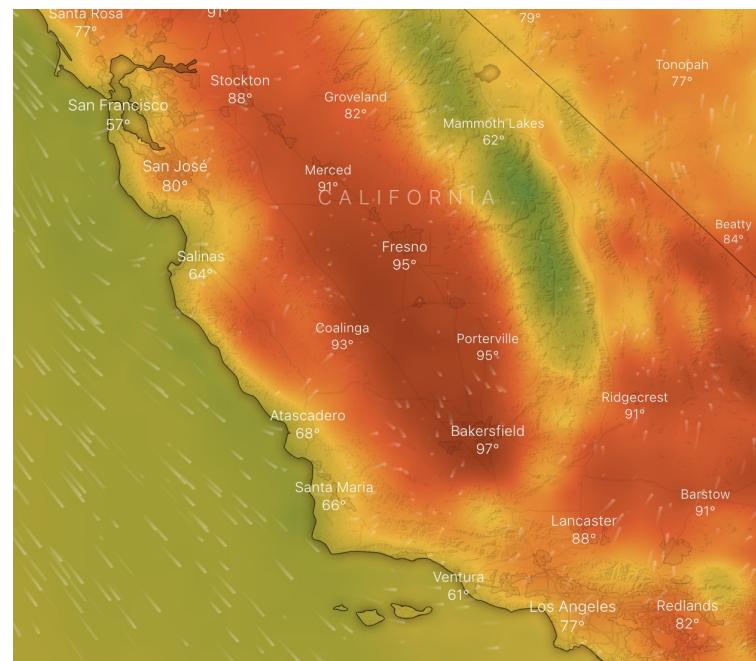
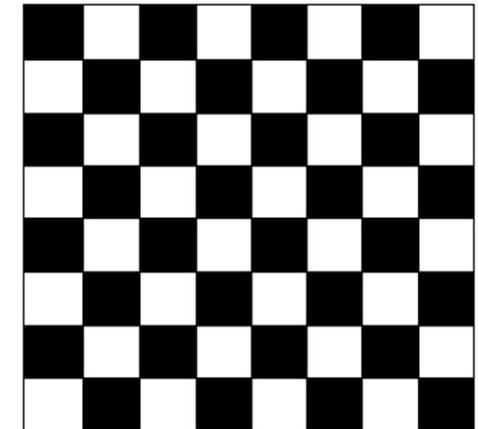
Spatial autocorrelation

Tobler's first law of geography: Everything is related to everything else.
But near things are more related than distant things.

Positive



Negative



Correlation and Moran's I

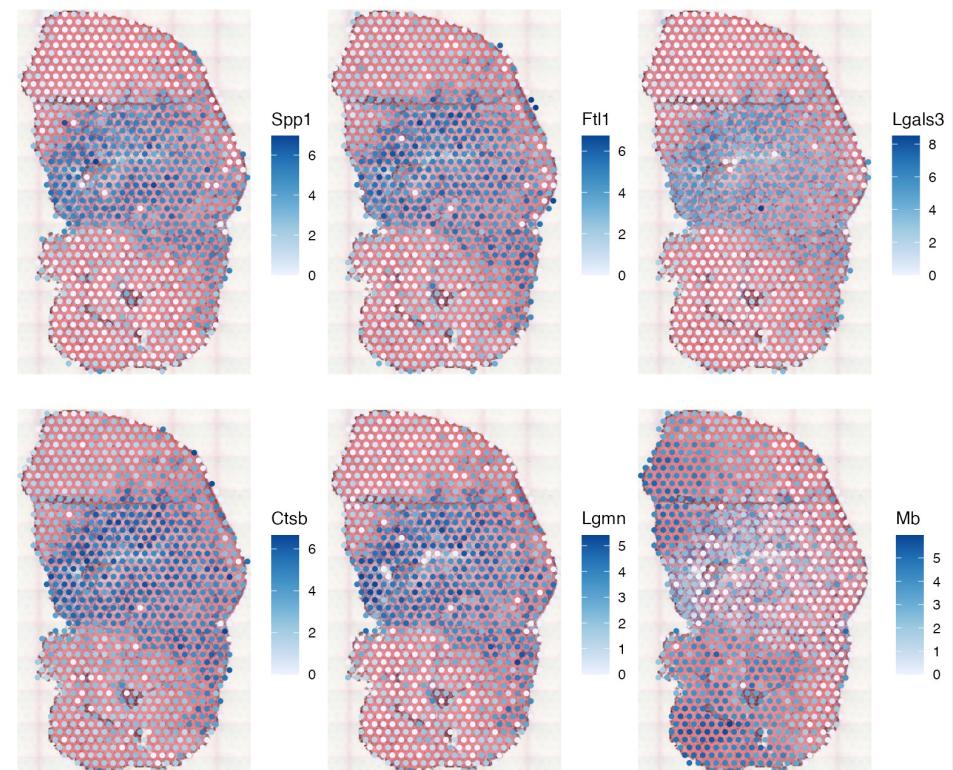
Pearson correlation: coexpression of 2 genes

$$r_{xy} = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2} \sqrt{\sum_{i=1}^n (y_i - \bar{y})^2}}$$

Moran's I: spatial autocorrelation of 1 gene

$$I = \frac{n}{\sum_{i=1}^n \sum_{j=1}^n w_{ij}} \frac{\sum_{i=1}^n \sum_{j=1}^n w_{ij}(x_i - \bar{x})(x_j - \bar{x})}{\sum_{i=1}^n (x_i - \bar{x})^2}$$

w_{ij} : spatial weights indicating neighborhood



Local Moran

w_{ij} : spatial weights

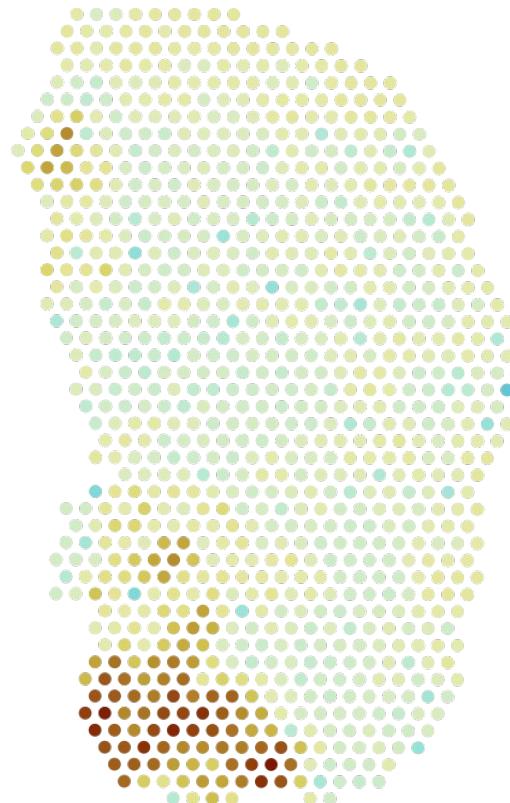
$$I = \frac{n}{\sum_{i=1}^n \sum_{j=1}^n w_{ij}} \frac{\sum_{i=1}^n \sum_{j=1}^n w_{ij} (x_i - \bar{x})(x_j - \bar{x})}{\sum_{i=1}^n (x_i - \bar{x})^2}$$

Local Moran:

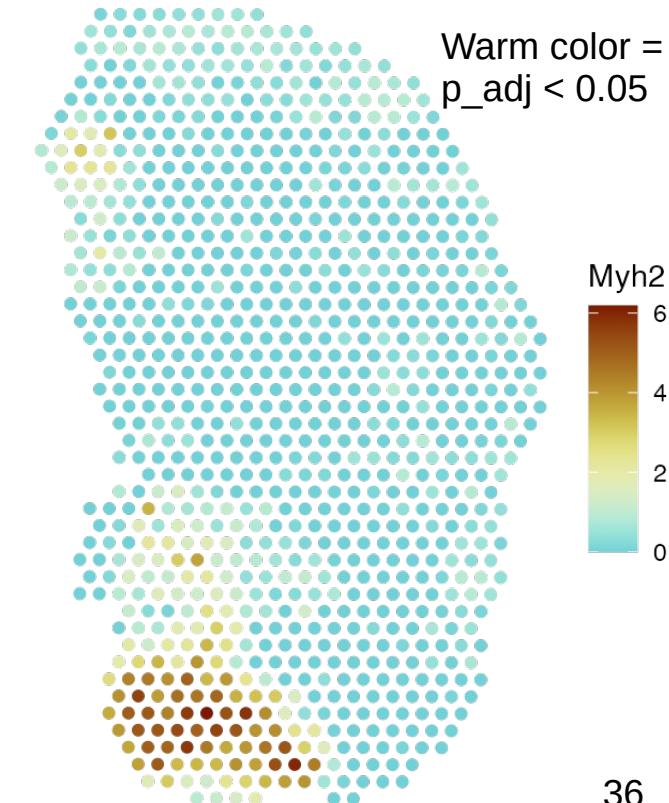
$$I_i = (n-1) \frac{(x_i - \bar{x}) \sum_{j=1}^n w_{ij} (x_j - \bar{x})}{\sum_{i=1}^n (x_i - \bar{x})^2}$$

Used in MERINGUE to identify spatially variable genes

Local Moran's I (li)



Local Moran's I (-log10p_adj)

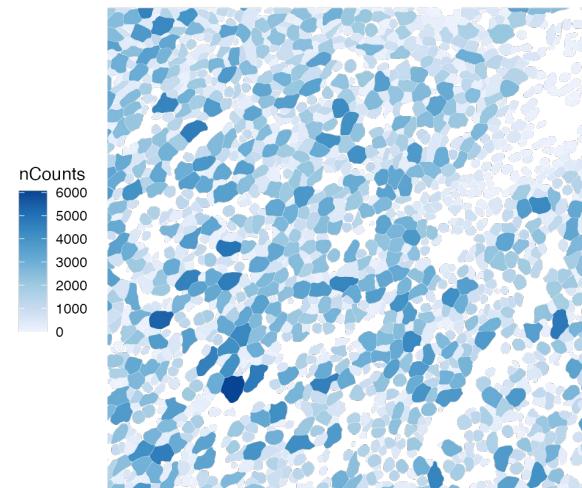
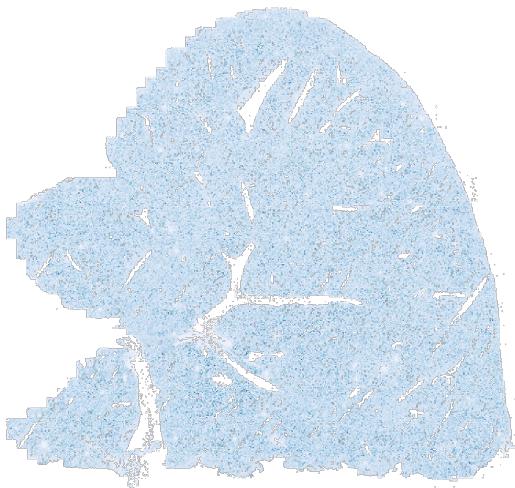


Scale matters

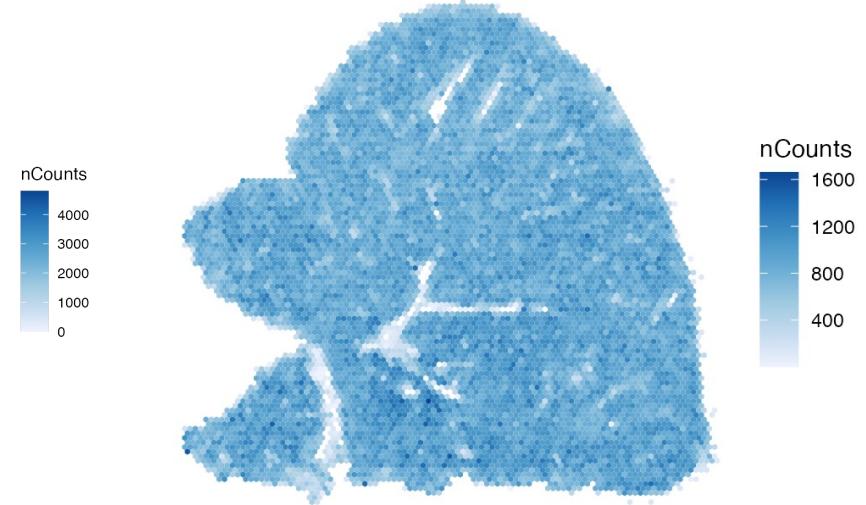


Positive and negative spatial autocorrelation are present at different scales

Single cell: Moran's I = -0.108

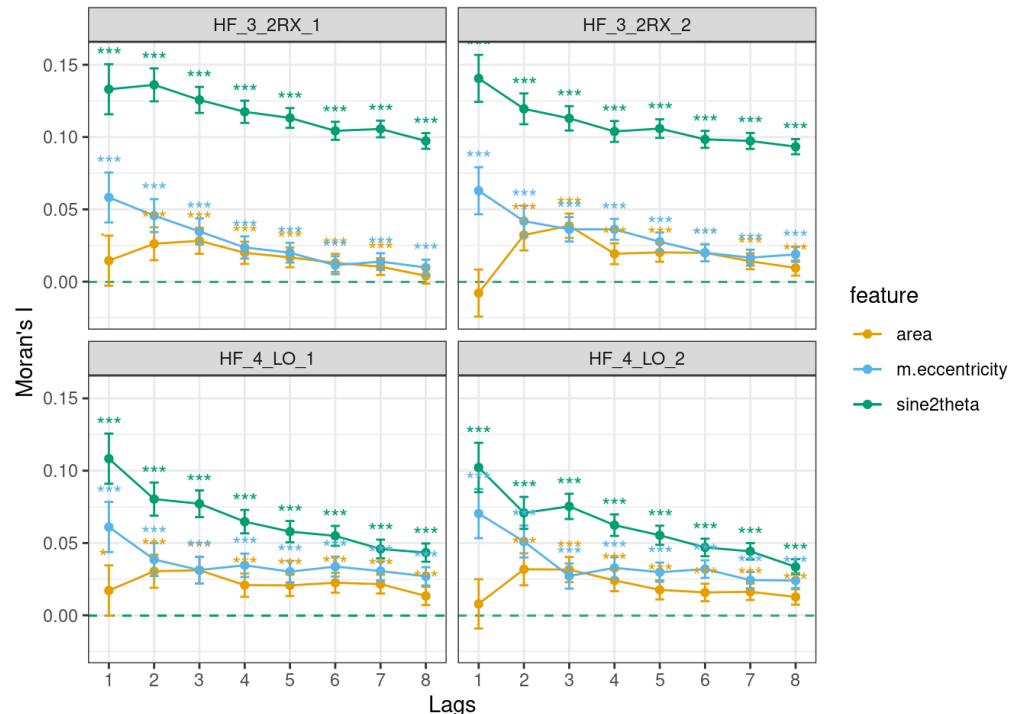


Binned: Moran's I = 0.491



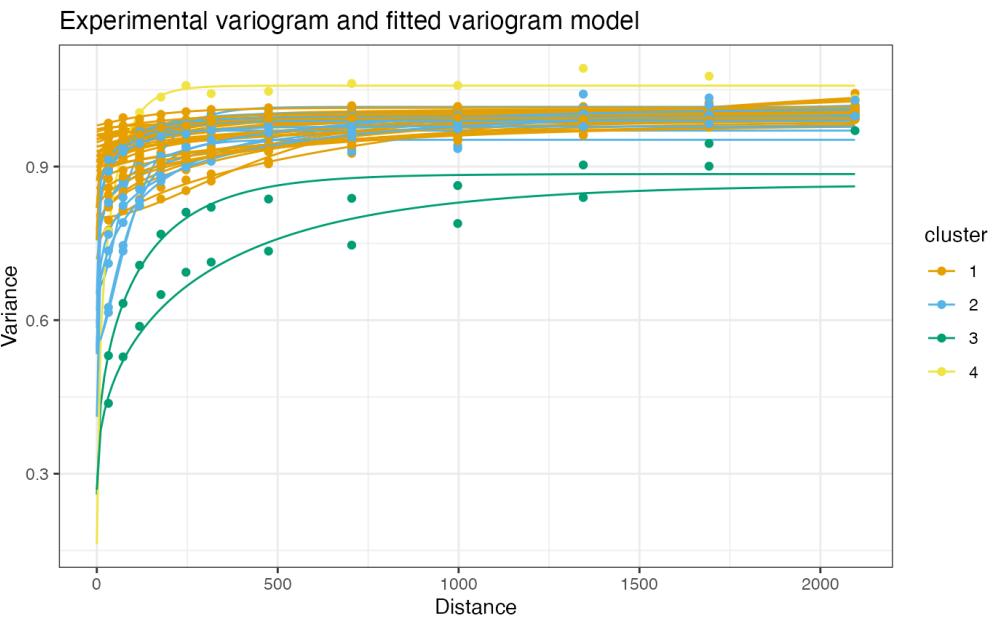
Length scale ESDA

Correlogram



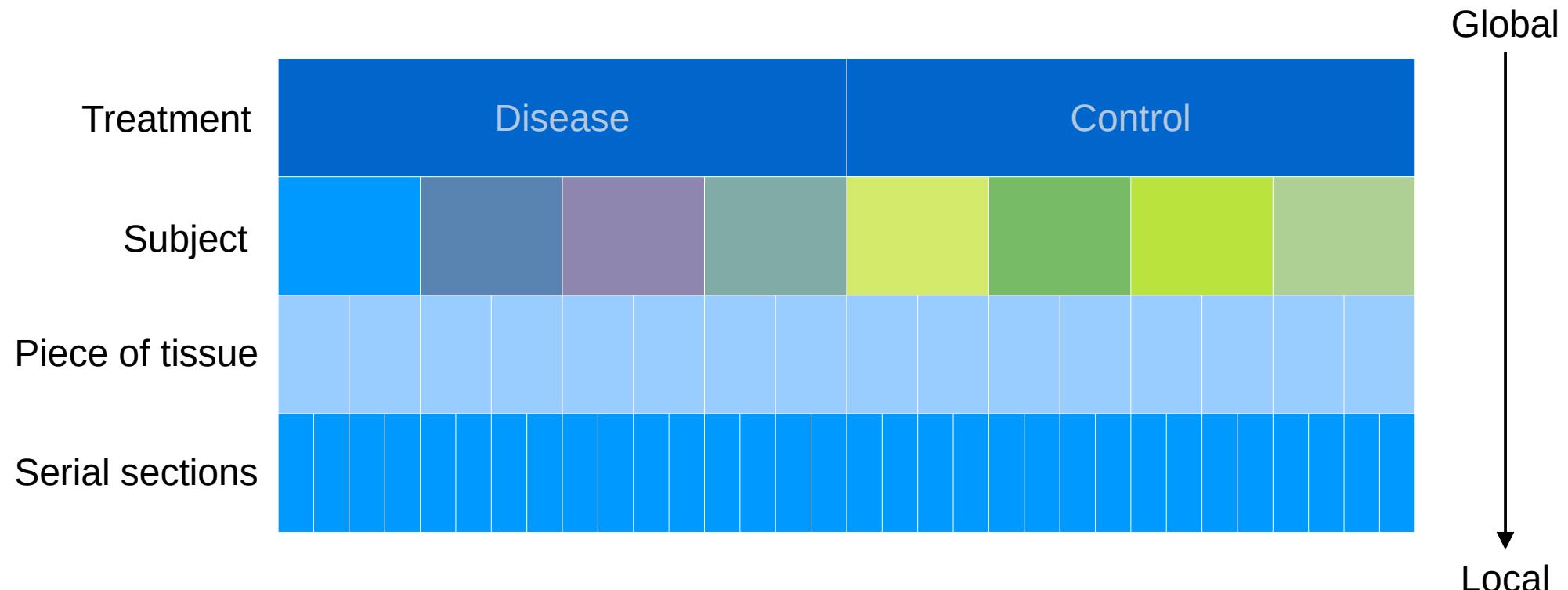
Adipocyte morphological metrics,
unpublished, from Marcus Seldin lab at UCI

Variogram

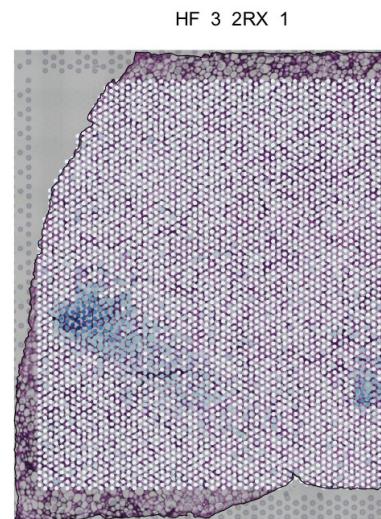
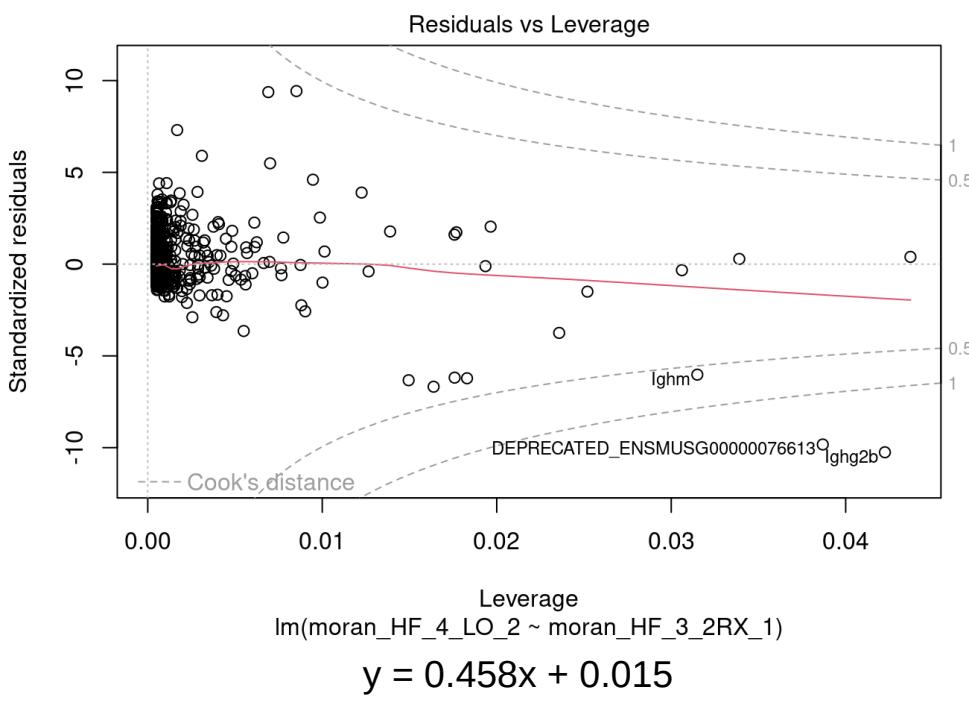


Biermann melanoma metastasis Slide-seq data

Hierarchy of sample relations, spectrum from global to local

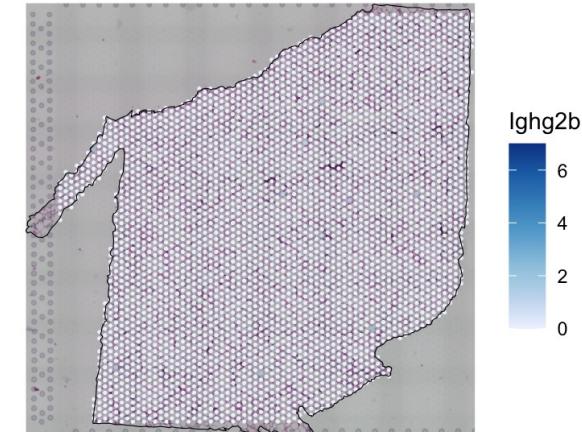


Comparing Moran's I in different samples



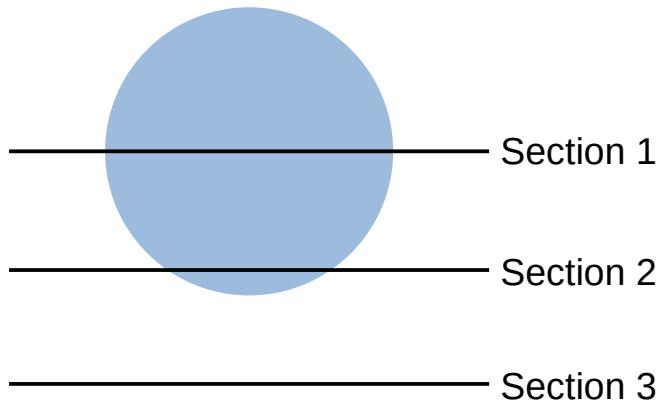
Control

High fat diet, 1 mouse/condition, 2 serial sections

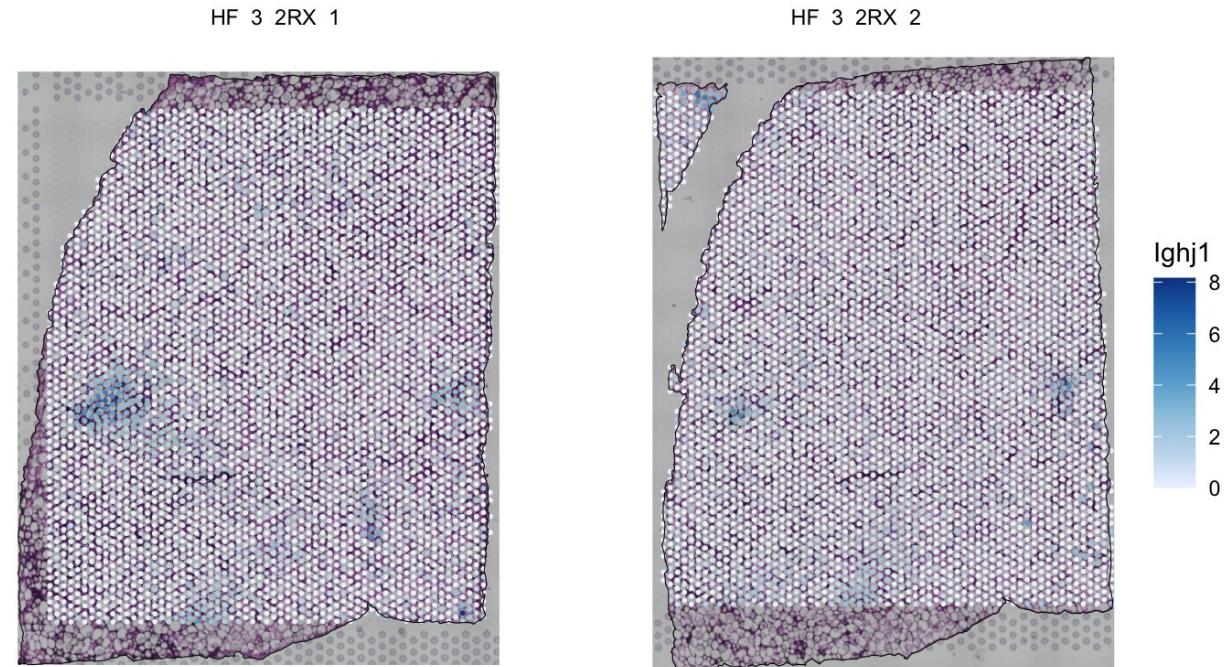


Itih5 KO

Experimental design problem



What if the section was just unlucky to miss the immune foci in the KO mouse?



Also see Three-dimensional assessments are necessary to determine the true, spatially-resolved composition of tissues, Forjaz et al. 2024

Future directions

- More on ESDA:
 - GWPCA, MSPA, spatial point process analysis of transcript spots
- From what I can do to doing it well:
 - Systematic examination of negative spatial autocorrelation in different tissues
 - Rigorous comparisons of spatial phenomena across samples and biological conditions, accounting for experimental design
 - Statistical method to associate gene expression to spatial covariates
 - Implications of data normalization and spatial neighborhood graphs

Acknowledgement

- Voyager website: Kayla Jackson, Laura Luebbert, Lior Pachter
- smFISH functionalities: Alik Huseynov (DKFZ)
- Tissue alignment: Cathal King (Univ. Adelaide)
- Python implementation: Sina Booeshaghi, Pétur Helgi Einarsson, Pall Melsted, Sindri Antonsson
- Joe Rich, Charlene Kim, and everyone in the lab
- UCI: Marcus Seldin, Mingqi Zhou
- Columbia: Bianca, Claire, Aga
- Bioconductor community: Alex Mahmoud, Hervé Pagès, Kasper Hansen, Allan O'Callaghan, Aaron Lun, Vincent Carey, Henrik Bengtsson, Erica Feick, Maria Doyle
- Funding: National Institute of Mental Health (NIMH), NIH, Icelandic Research Fund, Project, Health + Life Science Alliance Heidelberg Mannheim

