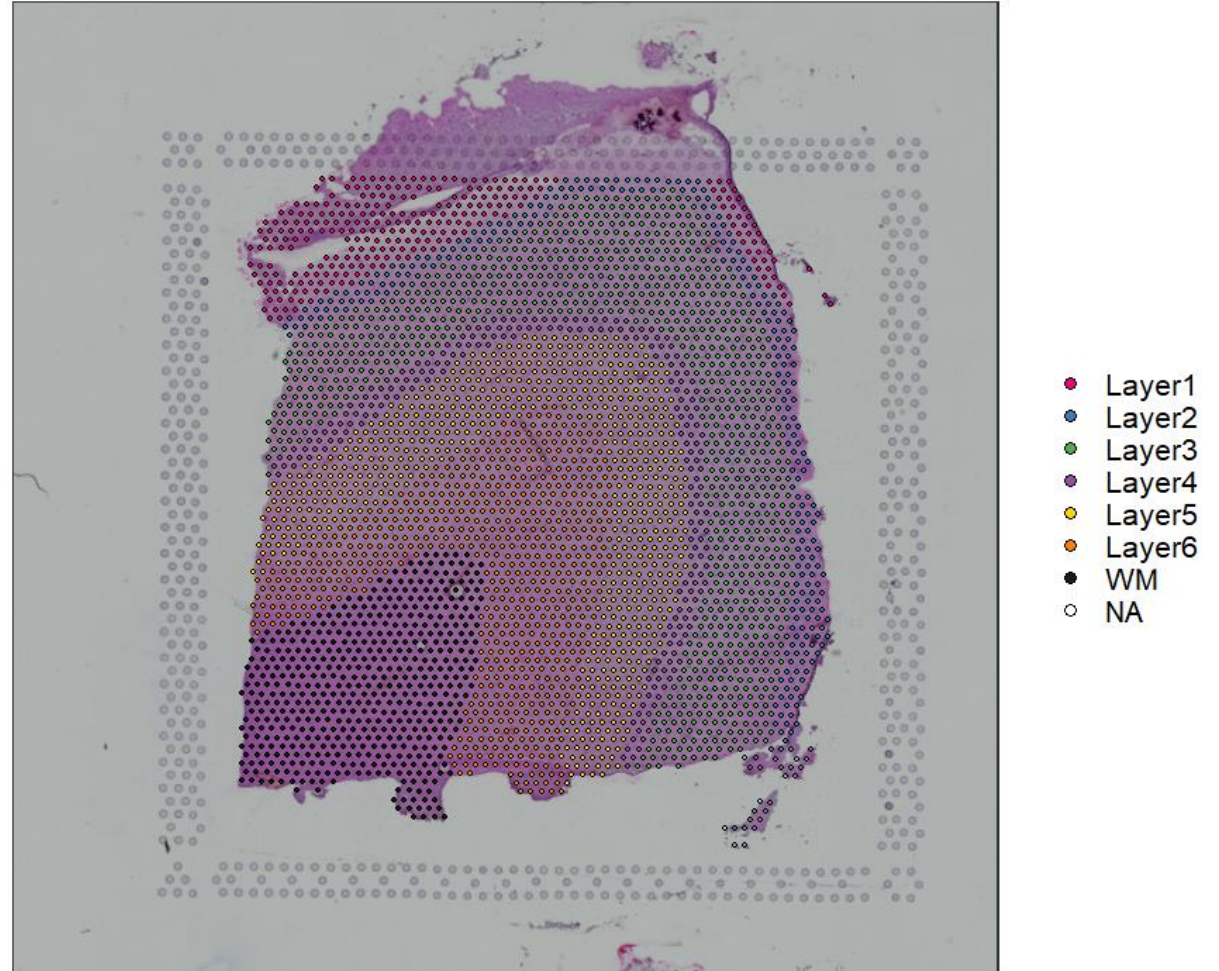


# Analysis of Maynard 2020 brain spatial transcriptomic data

4/6

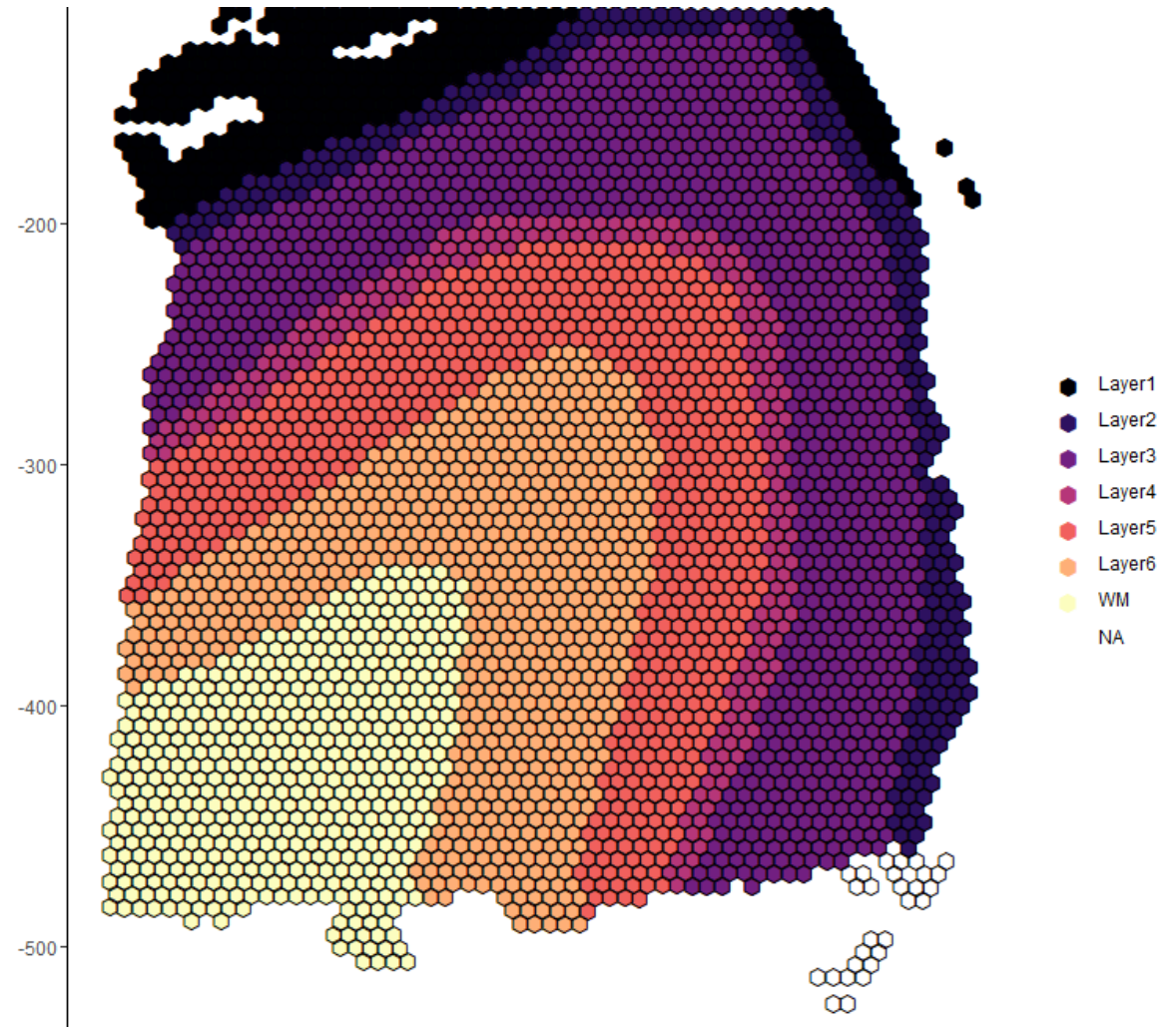
# Sample 151673

- Dorsolateral prefrontal cortex layers
- 3639 capture locations



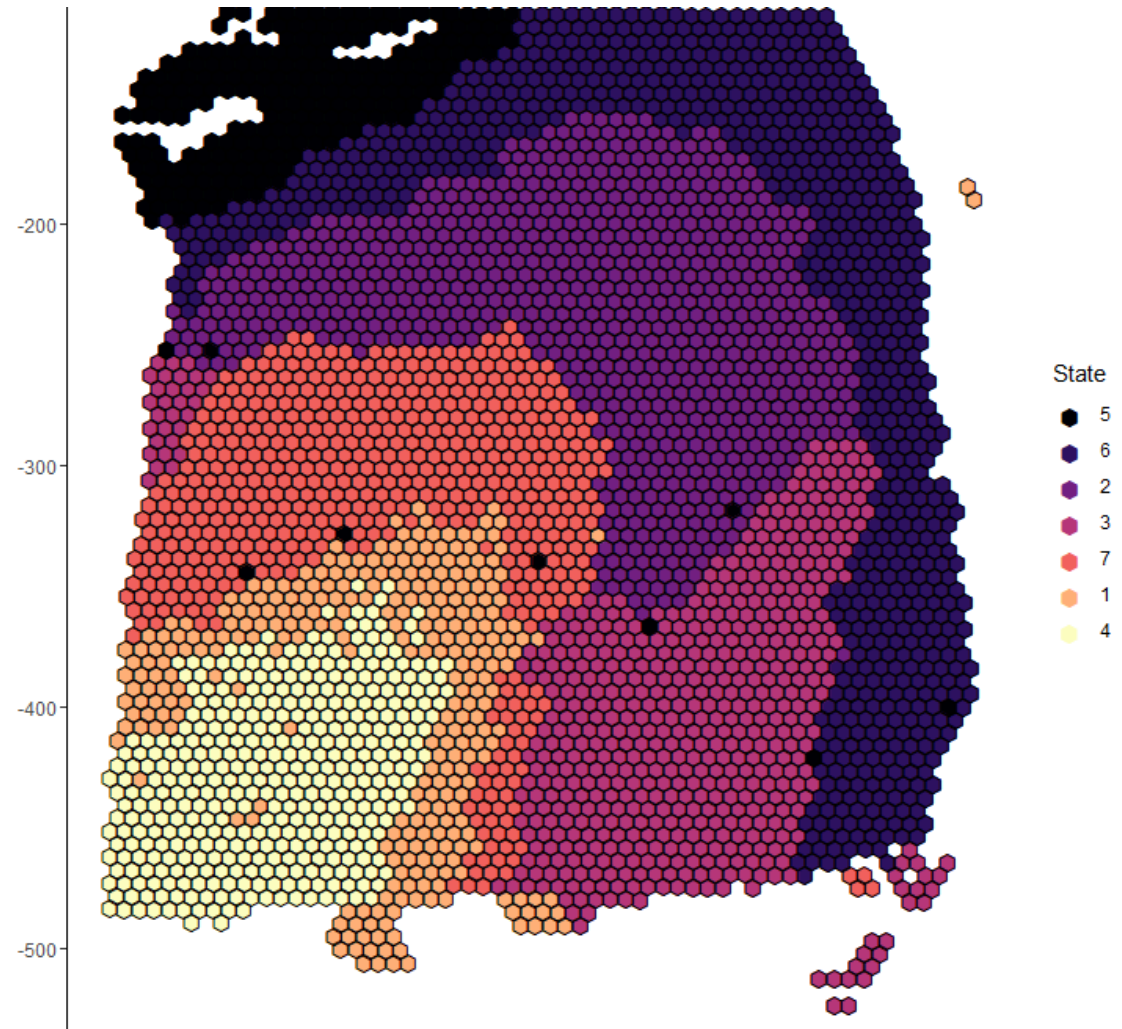
# Ground truth

- Supervised annotation based on “cytoarchitecture and selected gene markers”
- Clustering performance (ARI)
  - PCA on highly variable genes: 0.274
  - PCA on known layer makers: 0.197
  - PCA on differentially expressed genes between true layers: 0.380



# Spatial Clustering

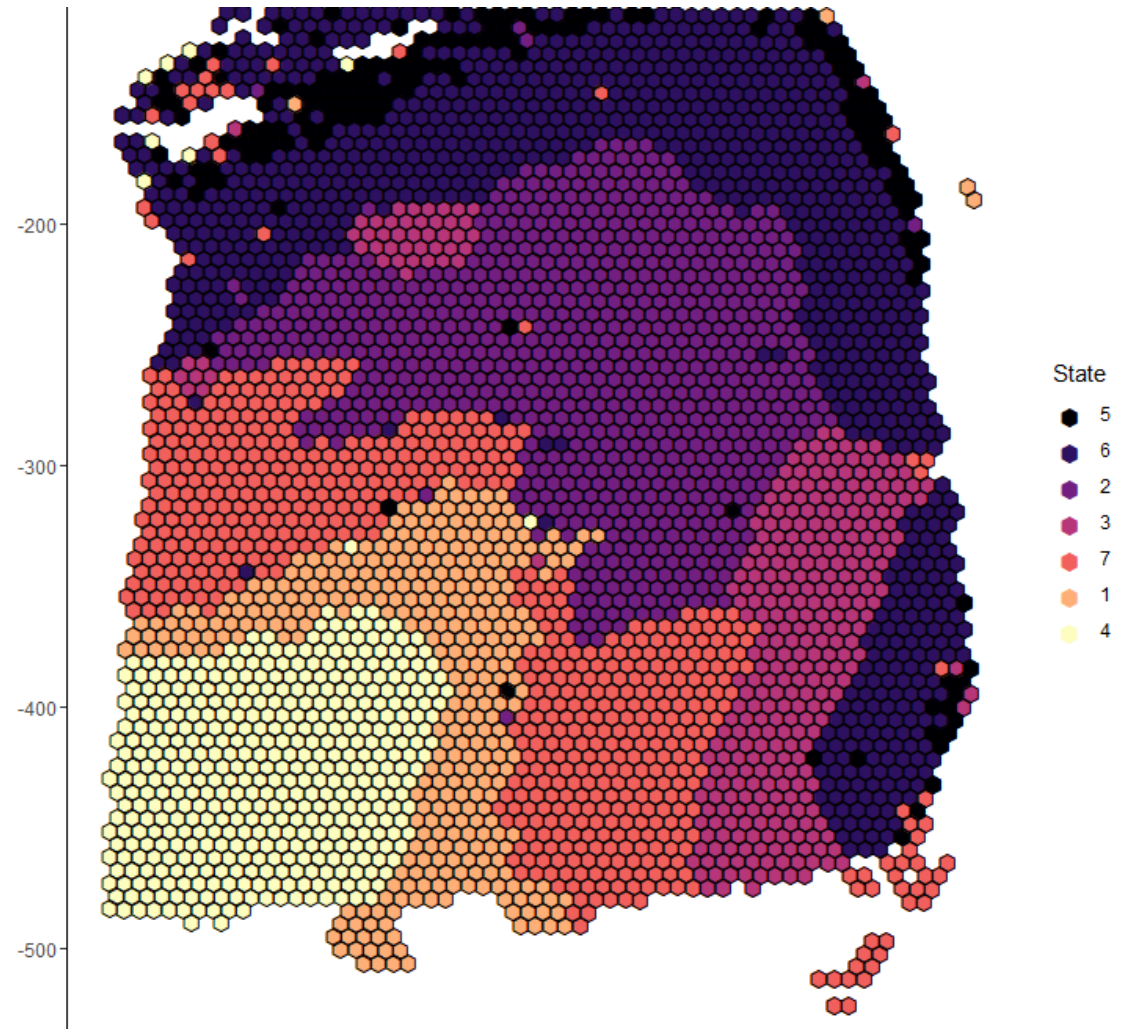
- 7 clusters
- 9 PCs
- ARI = 0.308





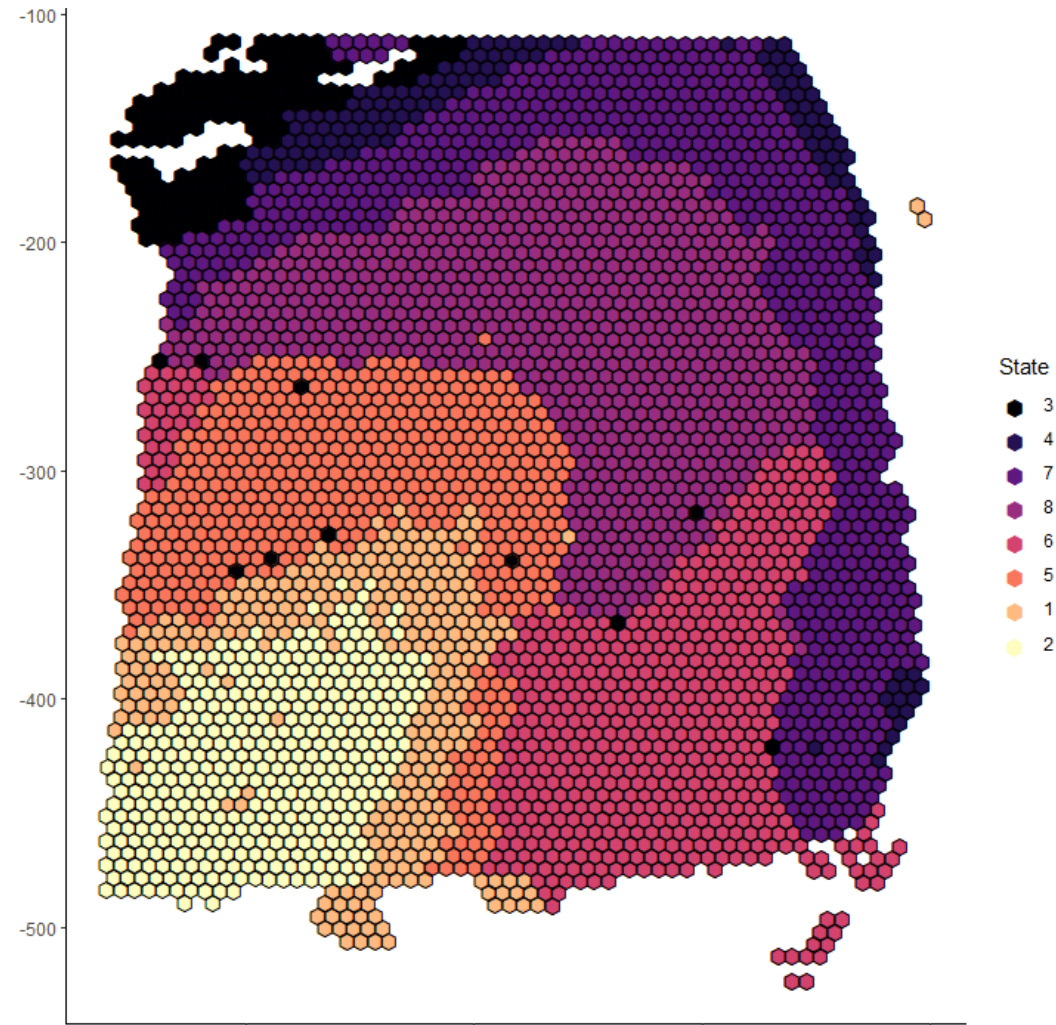
# Spatial Clustering

- 7 clusters
- UMAP
- ARI = 0.322



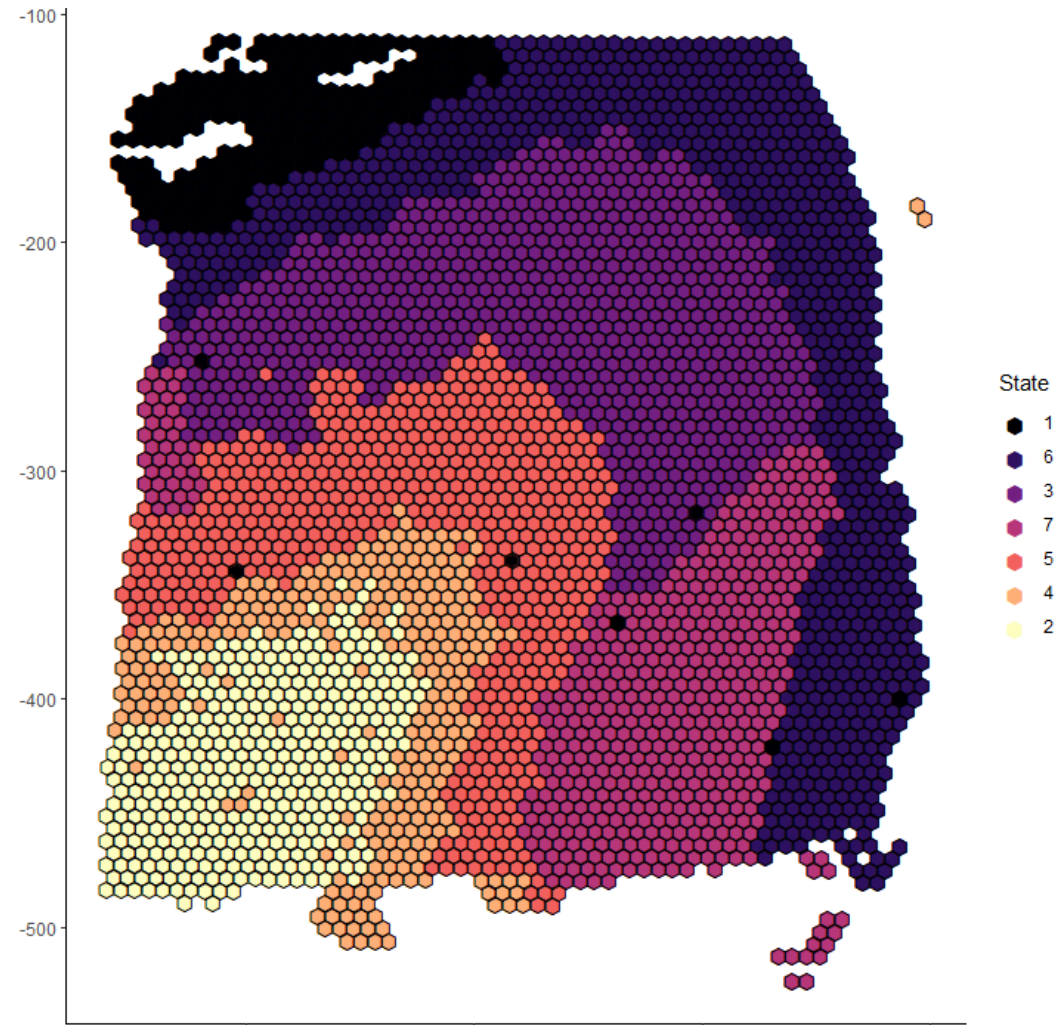
# Spatial Clustering

- 7 clusters
- t-SNE
- ARI = 0.332



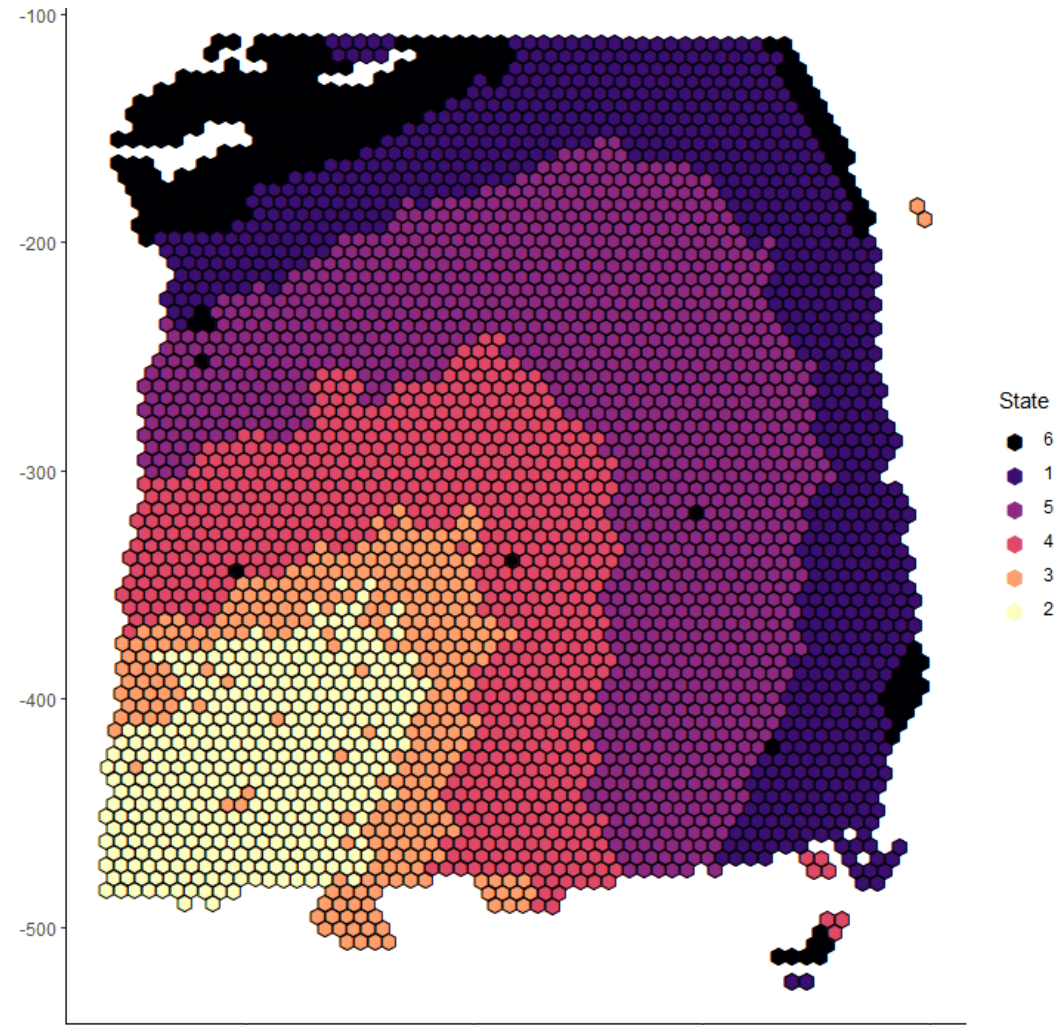
# Spatial Clustering

- 7 clusters
- 9 PCs using single sample
- ARI = 0.352



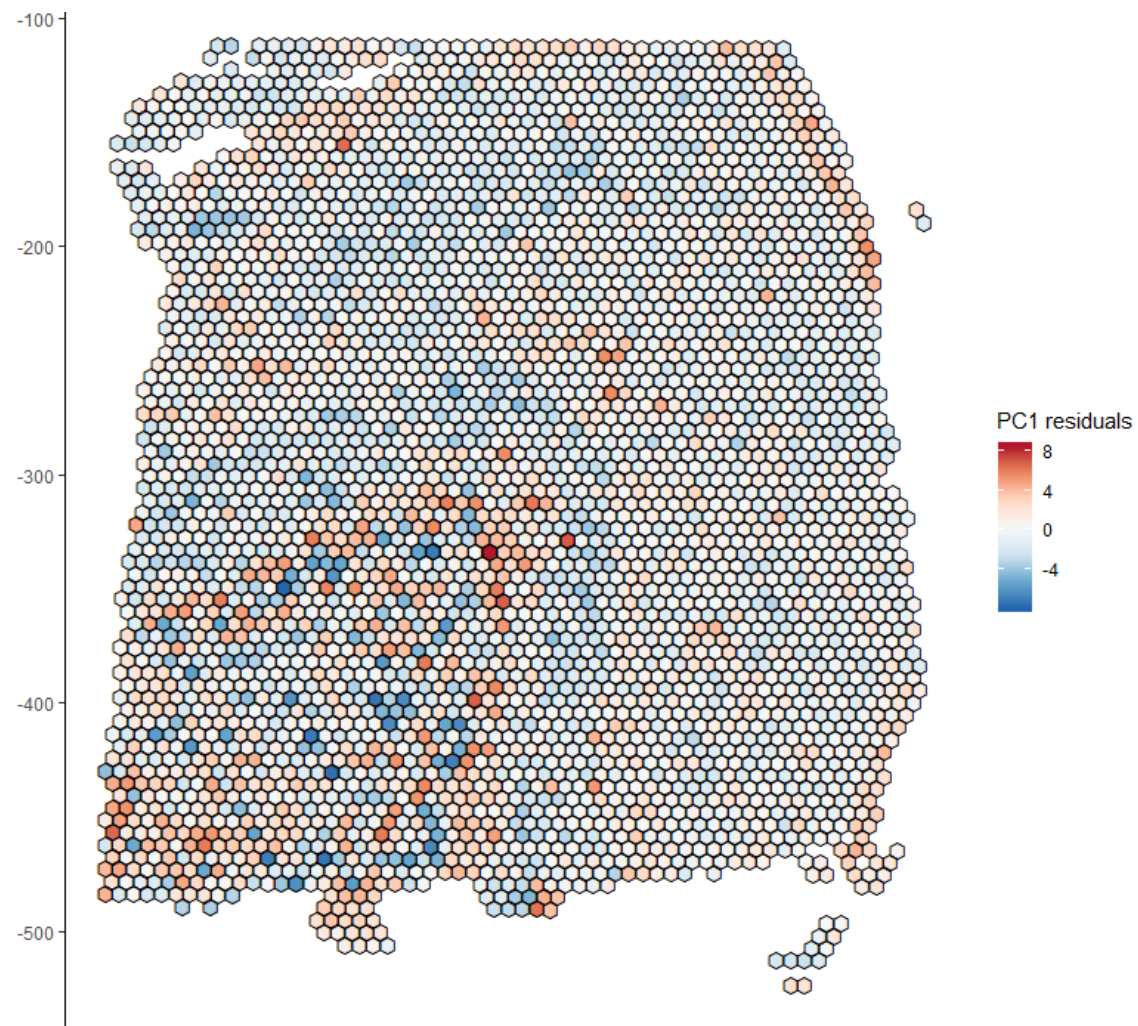
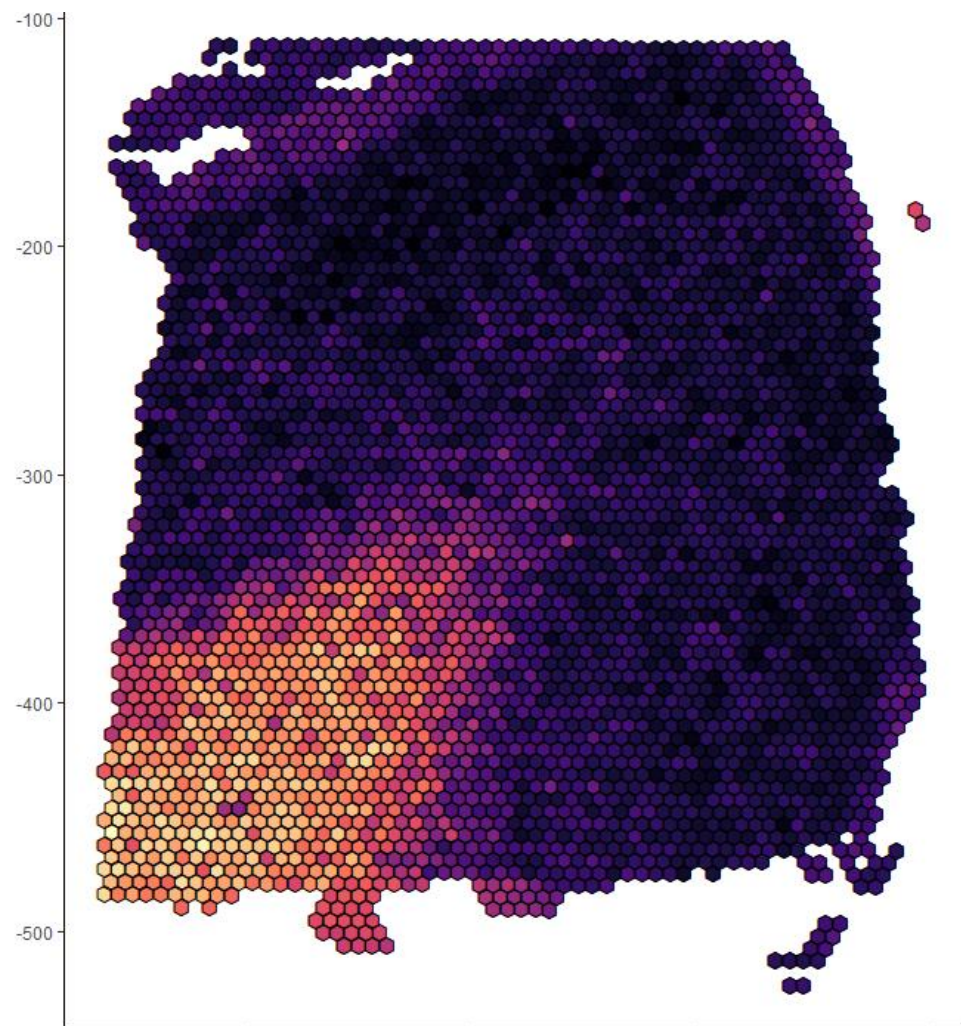
# Spatial Clustering

- 6 clusters
- 9 PCs using single sample
- ARI = 0.403

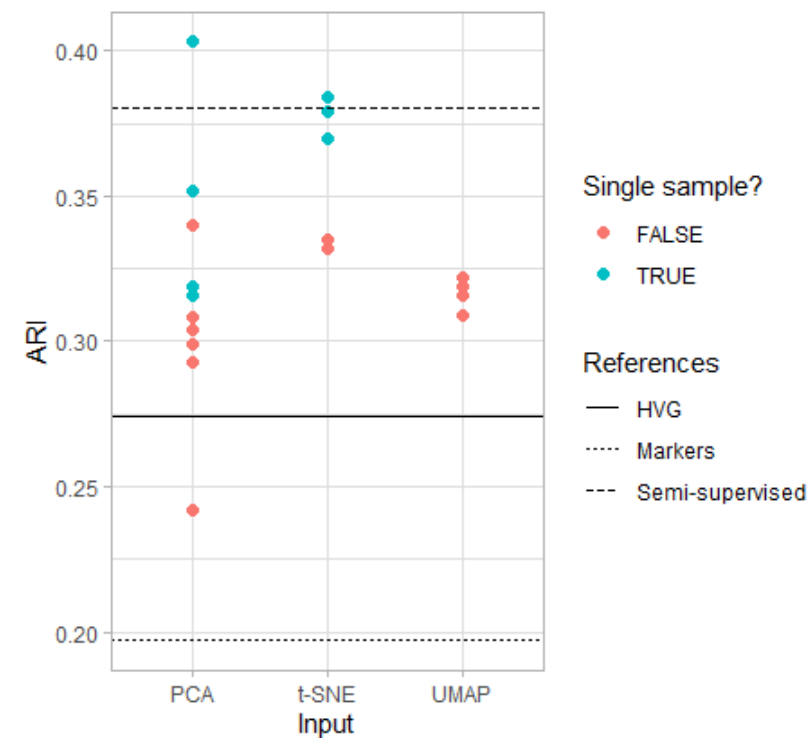
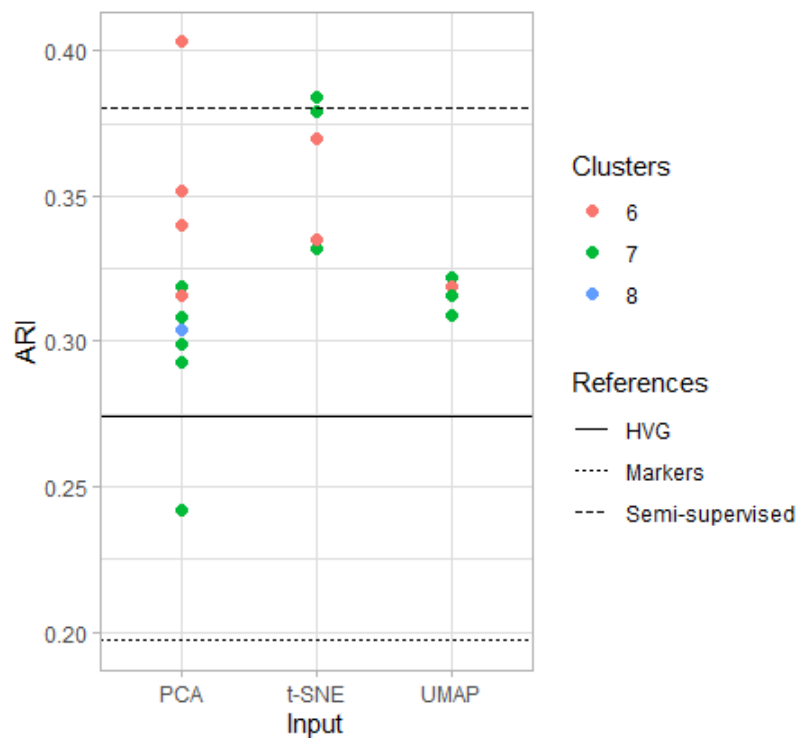
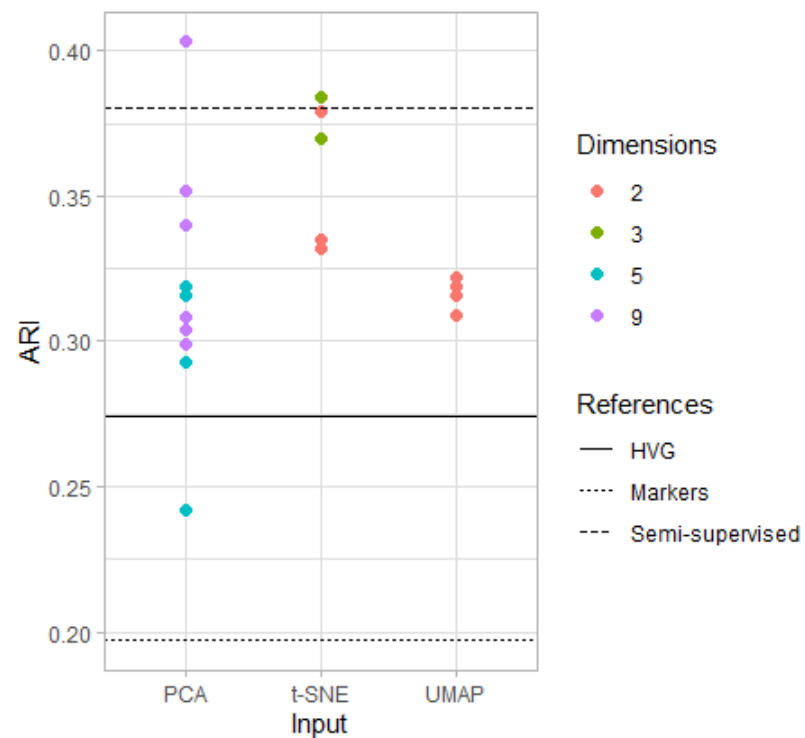




# PC1 vs. PC1 residuals



# Performance summary



# Additional thoughts

- Effect of initial seed (using UMAP input)
  - ARI between k means (2000 iter) vs. truth initialization (2000 iter): 0.566
  - ARI between k means (2000 iter) vs. truth initialization (20000 iter): 0.652
- Joint clustering of samples probably won't work well