

Background & Rationale

- Environmental science relies on multidimensional data
- Cluster & optimize data – Cosine & Euclidean
- How do you visualize relationships beyond a 3D graph?
- Visualization is crucial to all fields of research, particularly with environmental data
- It is important to understand which methods to use in different scenarios

Problem & Question

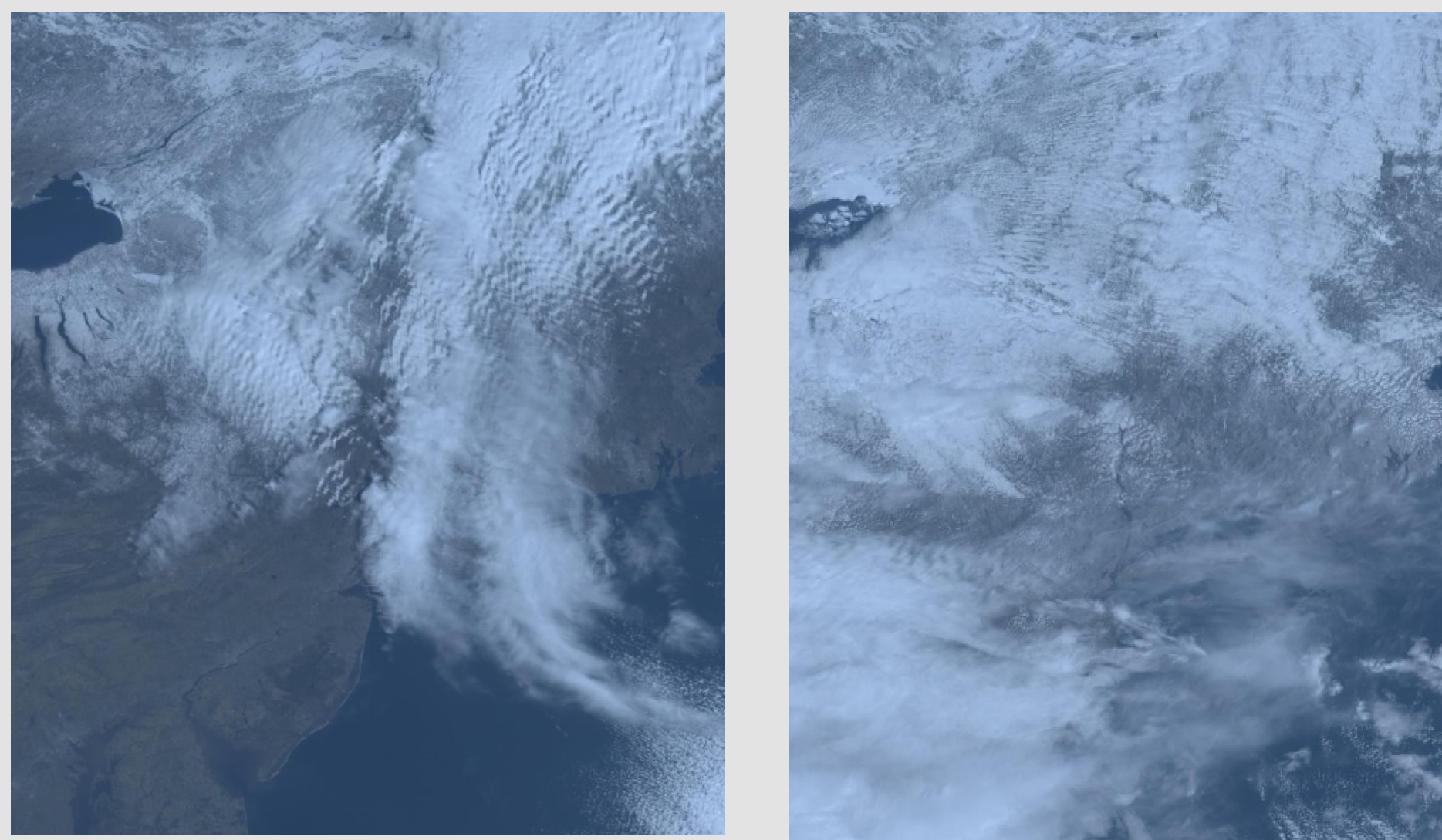
- There are a variety of clustering methods
- Which method is best to visualize?

Methods

- Tried different clustering tools
- Assessed in Java VCS
- First used 500 Dimensions, then repeated with 1000
- Used Altair to compare clustering type, number, and silhouette score

The Data

NOAA Satellite image data of 1000 dimensions



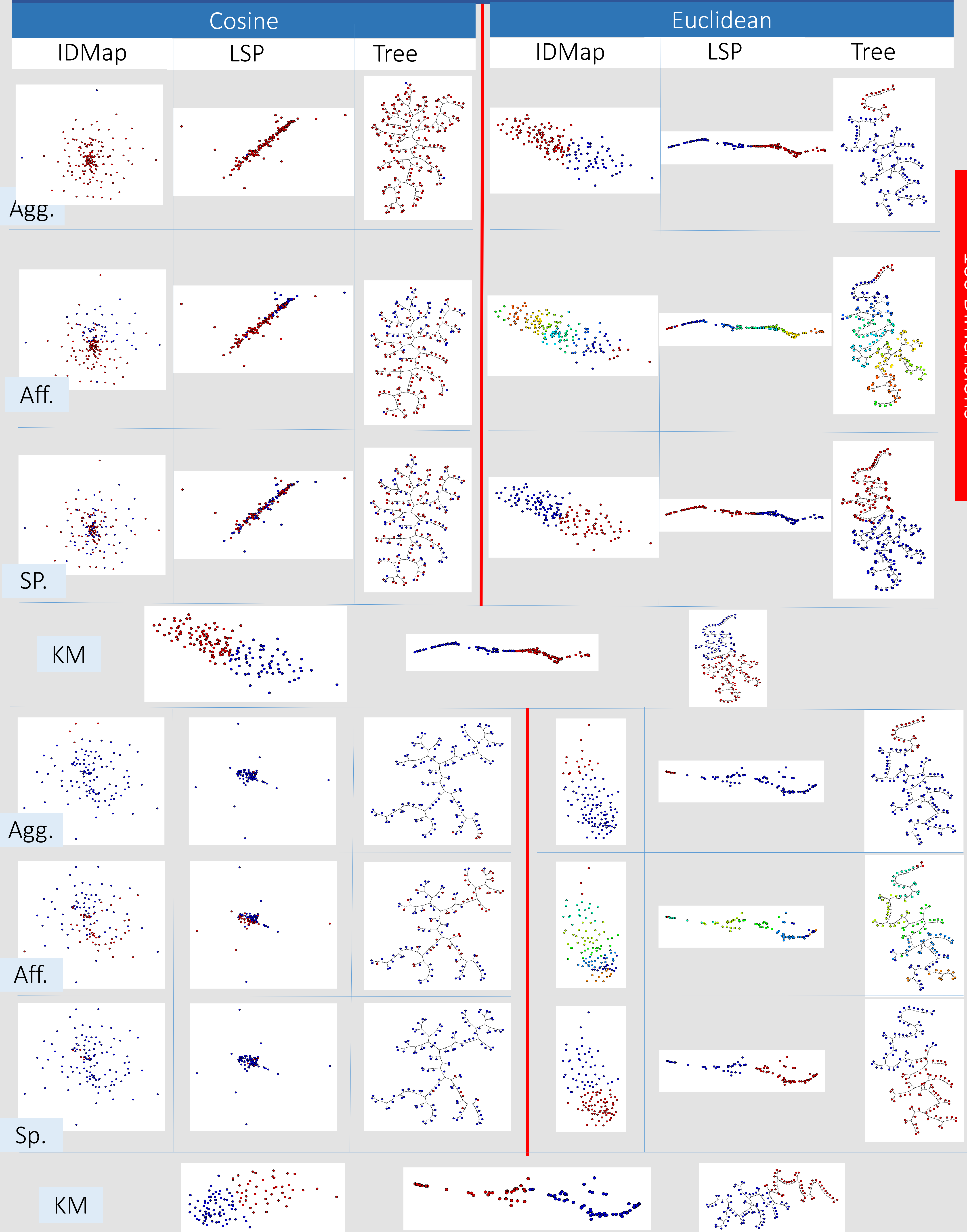
Conclusions

- Euclidean had better segregation
- The optimal number of clusters was 2
- Projection was more accurate than clustering
- Increasing the number of dimensions improved the clustering

References

- R. Etemadpour, R. Motta, J. G. d. S. Paiva, R. Minghim, M. C. F. de Oliveira and L. Linsen, "Perception-Based Evaluation of Projection Methods for Multidimensional Data Visualization," in *IEEE Transactions on Visualization and Computer Graphics*, vol. 21, no. 1, pp. 81-94, 1 Jan. 2015. doi: 10.1109/TVCG.2014.2330617

Results



100 Dimensions

1000 Dimensions

Silhouette Scores

