Inference, Computation, & Dynamic Visualization for Convex Clustering

Genevera I. Allen

Departments of Statistics, Computer Science,
and Electrical and Computer Engineering, Rice University,
Jan and Dan Duncan Neurological Research Institute,
Baylor College of Medicine & Texas Children's Hospital.

June 5, 2018

Motivation: Clustering & Biclustering

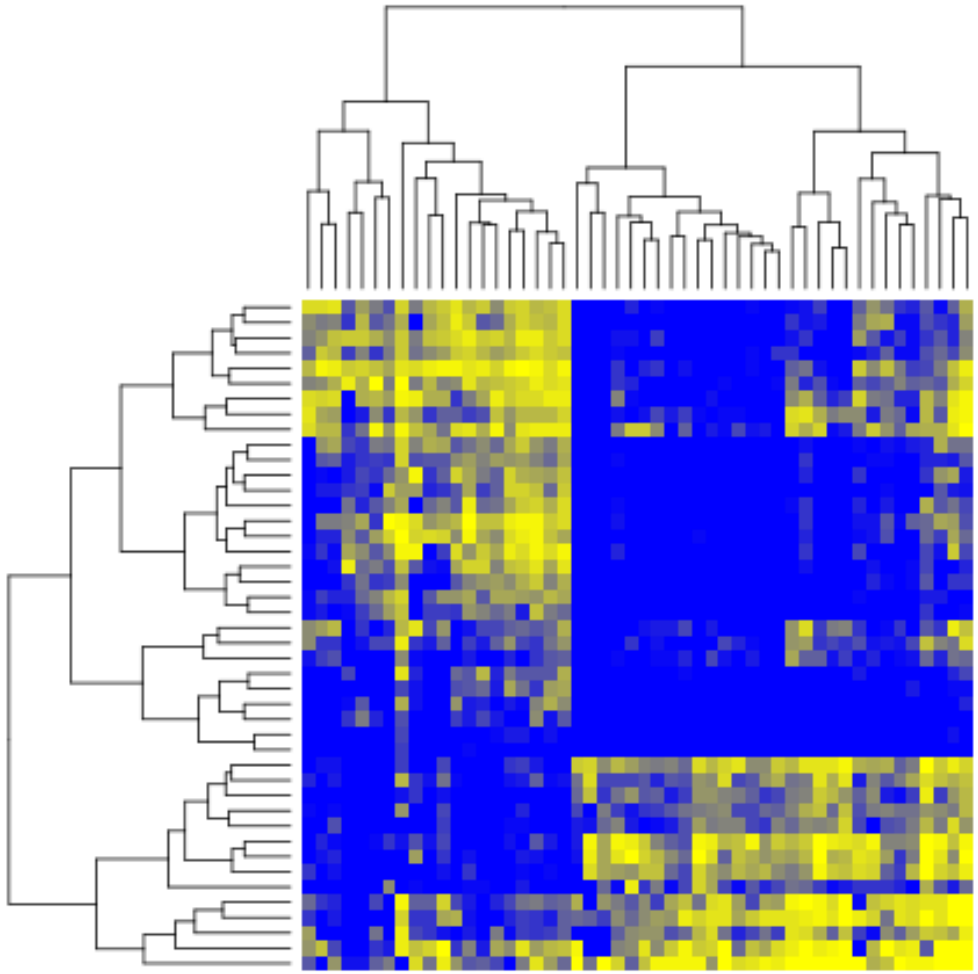
Clustering

Find groups of objects which are similar to each other.

Biclustering

Simultaneously find groups of features & observations.

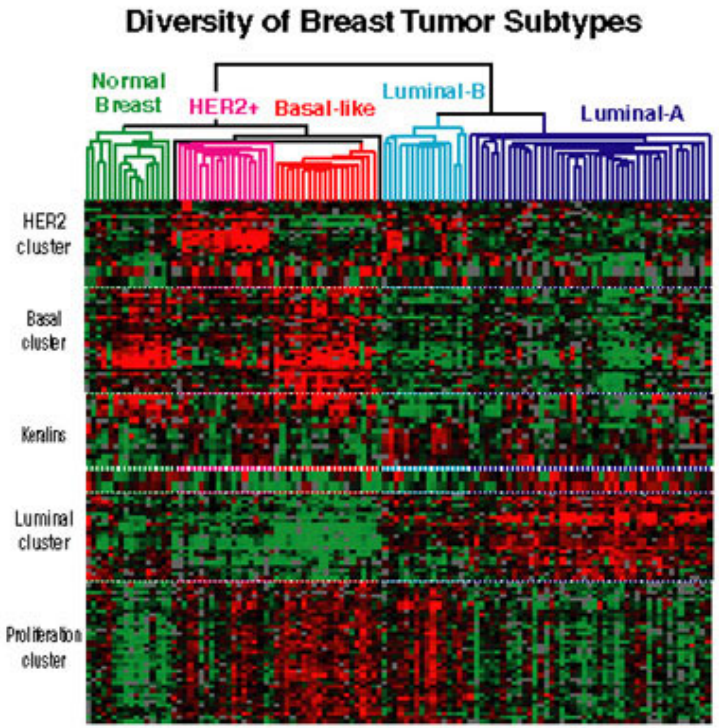
- Cluster rows & columns of data matrix.



Motivation: Clustering & Biclustering

Cancer Subtypes:
Groups of patients that share distinct genomic signatures and exhibit differing clinical outcomes.

- Major Success Story:
  Breast Cancer.
- Significance:
  Precision medicine!



Motivation: Clustering & Biclustering

Text Mining:

	data	R	big	cluster	shiny	fast	plot
doc 1	57	1	43	2	0	22	4
doc 2	17	29	2	3	35	6	44
doc 3	47	33	0	0	24	3	19
doc 4	23	0	0	31	0	7	2
doc 5	40	5	28	9	0	21	6
doc 6	8	10	7	46	12	17	9

Motivation: Clustering & Biclustering

Recommender Systems:

2			4	5	
5		4			1
		5		2	
	1		5		4
		4			2
4	5		1		

Clustering Approaches

The Good:

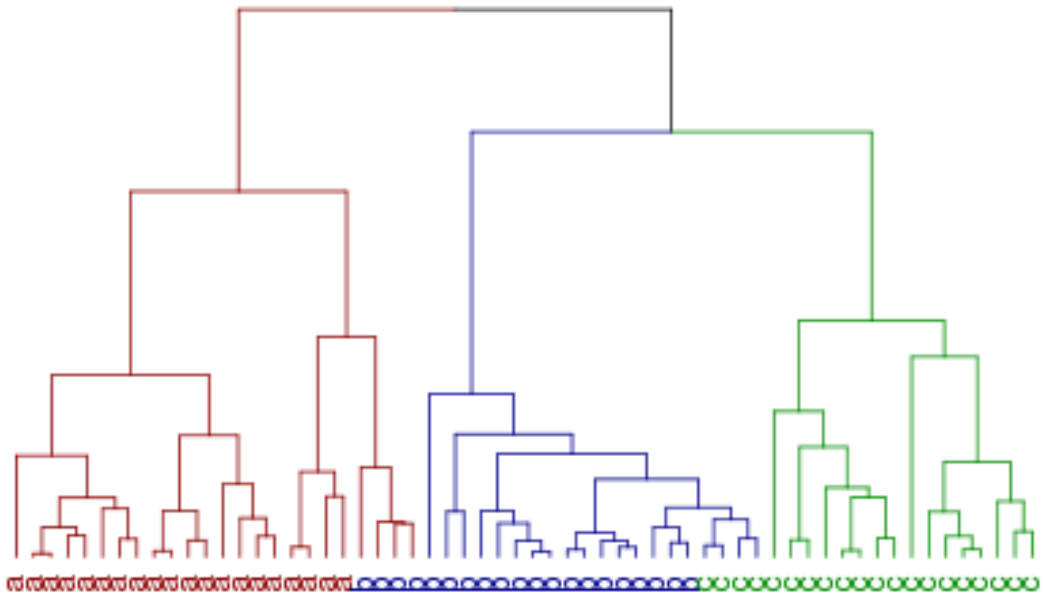
- Simple & Fast.
- Appealing Visualizations.
- Easy Interpretation.

The Bad:

- Local solutions.
- Instability.
- Tuning parameters.

The Ugly:

- How many clusters?
- Inference.



Hierarchical

Convex Clustering & Biclustering

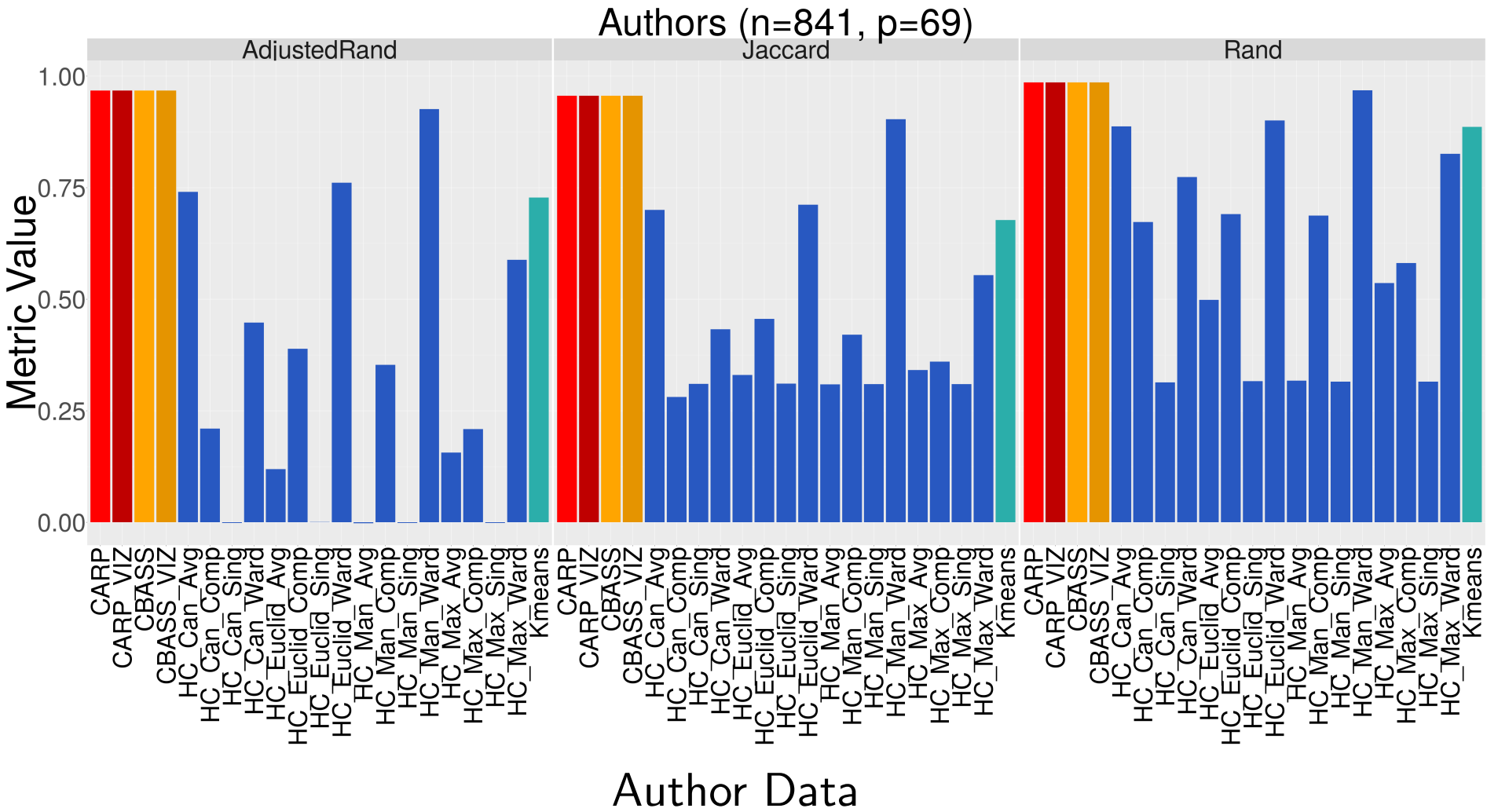
Solution: Convex Clustering & Biclustering

Why Convex?

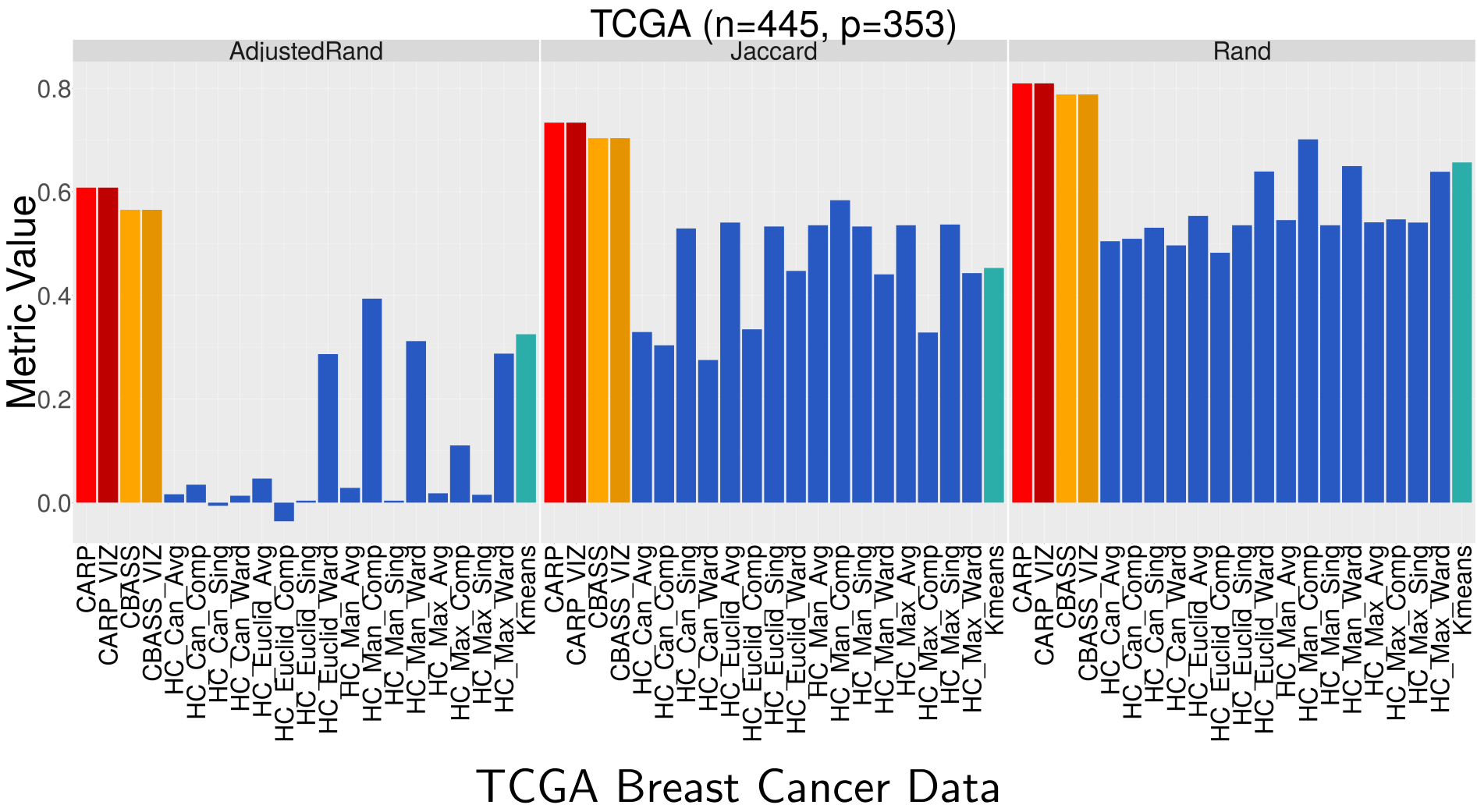
- Global solution!
- Superior mathematical and statistical properties:
  - Consistency.
  - Stability.
  - Improved clustering performance.
- Data-driven selection of # of clusters.

Pelckmans et al. 2005; Lindsten et al. 2011; Hocking et al. 2011; Chi & Lange 2013; Tan & Witten 2015; Chi, Allen & Baraniuk, 2017; Radchenko & Mukherjee, 2017

Clustering Accuracy



Clustering Accuracy



Convex Clustering

$$\underset{\mathbf{u}}{\text{minimize}} \quad \frac{1}{2}\sum_{i=1}^{n}\|\mathbf{x}_i - \mathbf{u}_i\|_2^2 + \lambda \sum_{i<j} w_{ij}\|\mathbf{u}_i - \mathbf{u}_j\|_2$$

- $\mathbf{x}_i$ - each observation (p-vector).
- $\mathbf{u}_i$ - cluster centroid for each observation.

Convex fusion penalty shrinks centroids together!

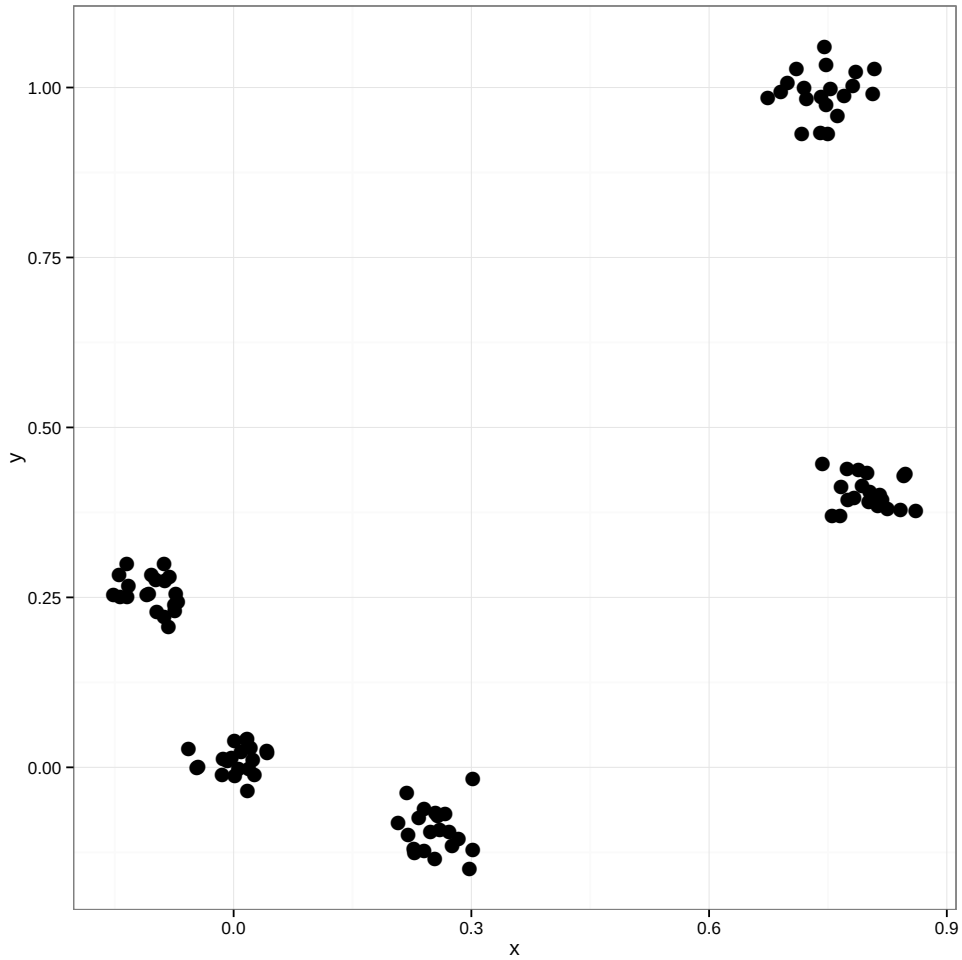
Pelckmans et al. 2005; Lindsten et al. 2011; Hocking et al. 2011; Chi & Lange 2013; Tan & Witten 2015.

Convex Clustering

$$\underset{\mathbf{u}}{\text{minimize}} \quad \frac{1}{2}\sum_{i=1}^{n}\|\mathbf{x}_i - \mathbf{u}_i\|_2^2 + \lambda \sum_{i<j} w_{ij}\|\mathbf{u}_i - \mathbf{u}_j\|_2$$

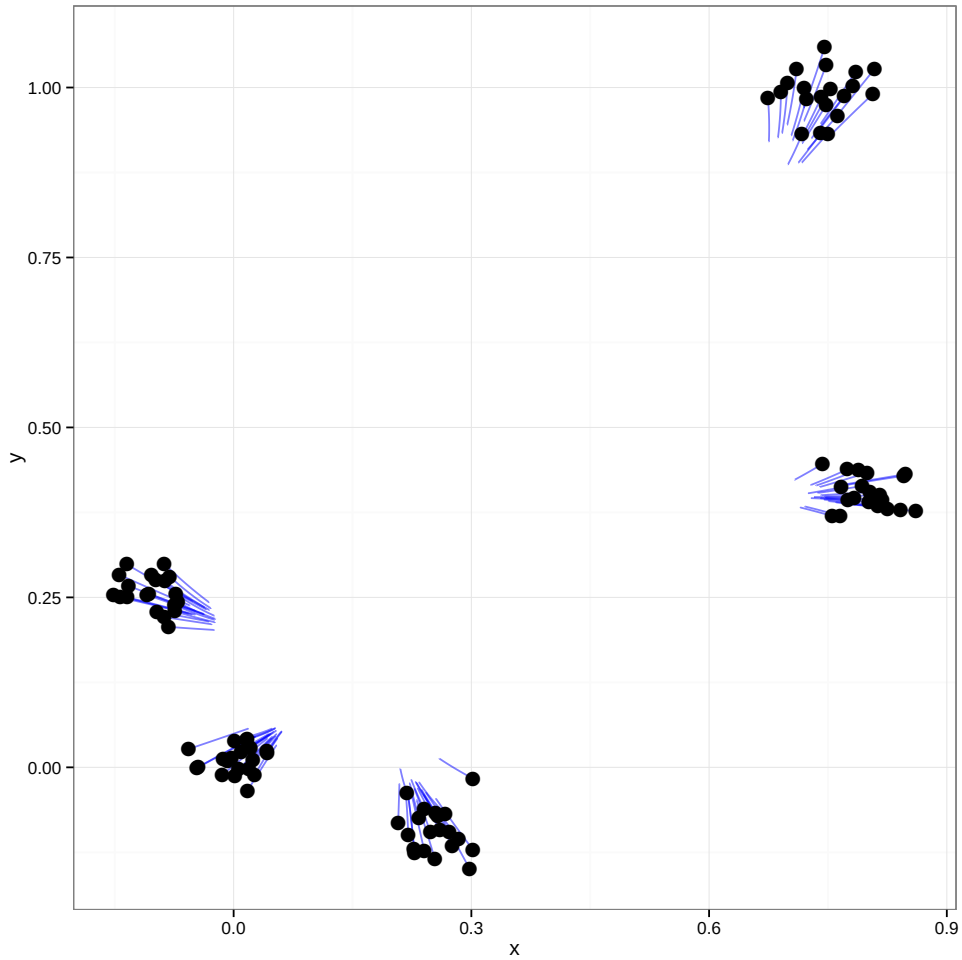
- λ controls BOTH cluster assignments & number of clusters.
  - $\lambda = 0$ - each observation is its own cluster.
  - λ larger - column means begin to coalesce together into clusters.
  - λ very large - all observations fused into one cluster.
- Algorithm: Alternating Minimization Algorithm.
- In R: `cvxclustr`.

Convex Clustering Solution Path



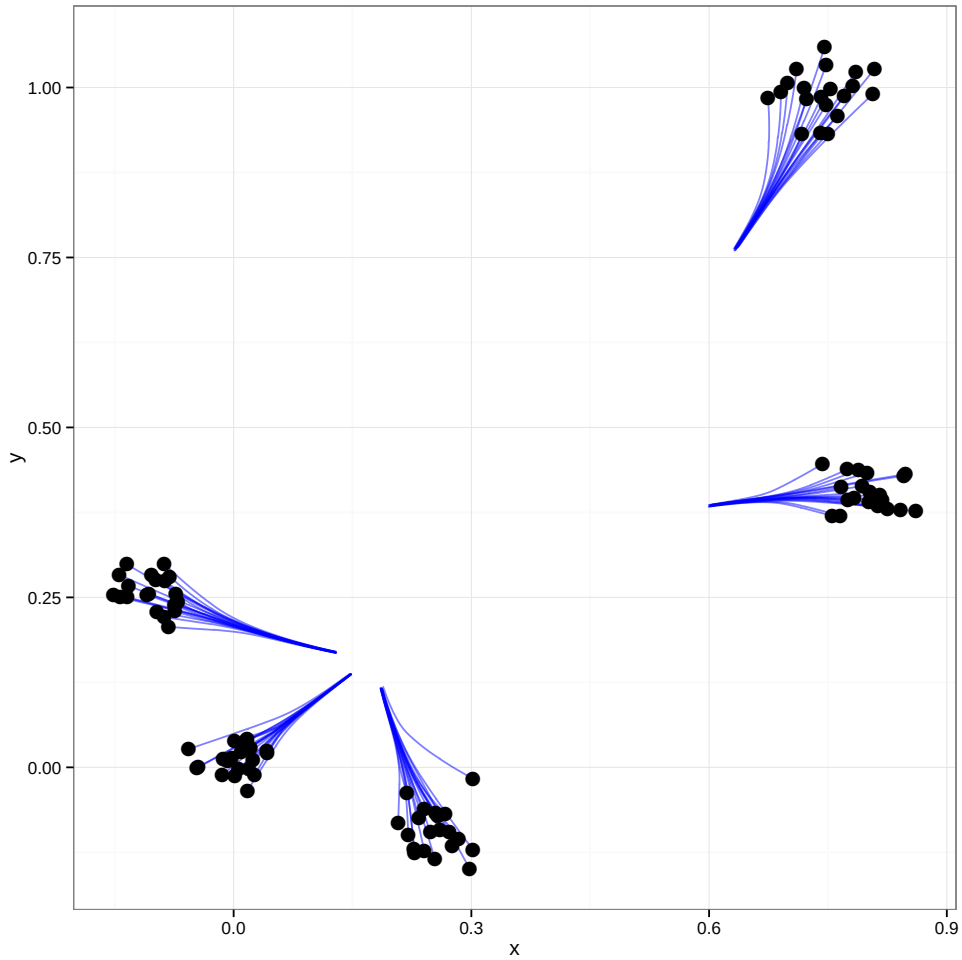
$\lambda = 0$

Convex Clustering Solution Path



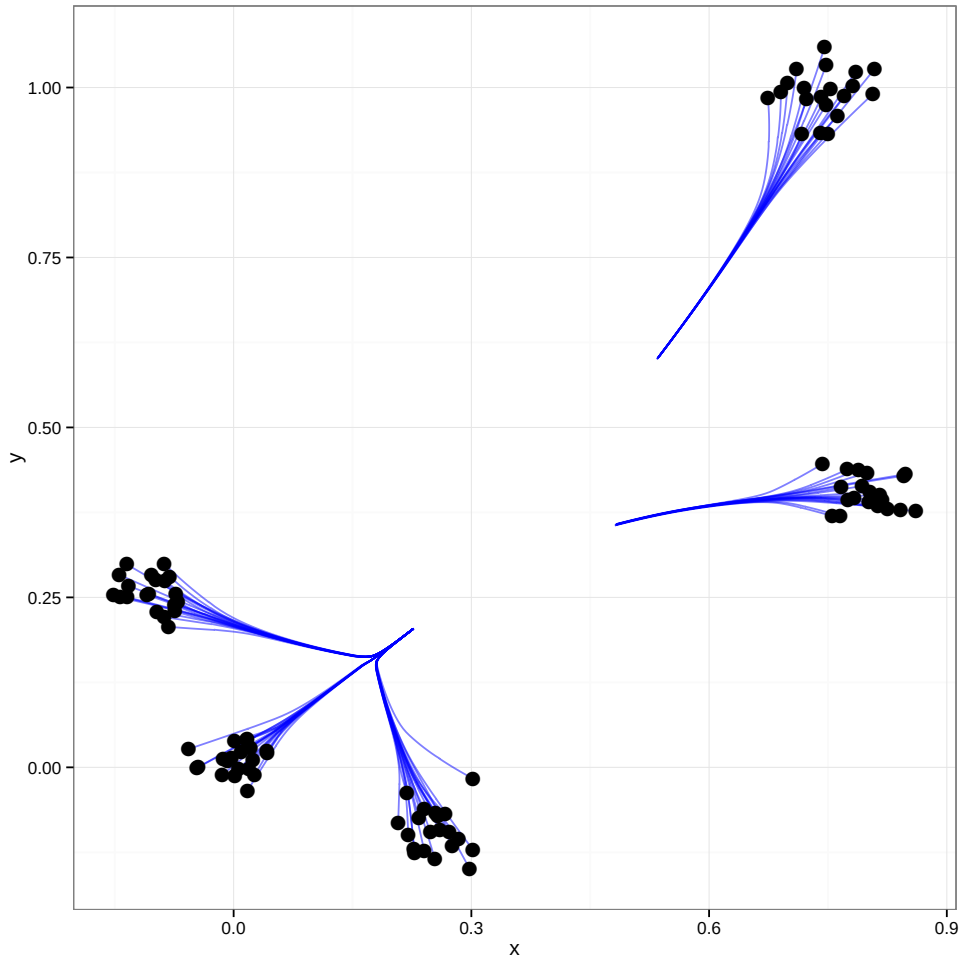
λ

Convex Clustering Solution Path



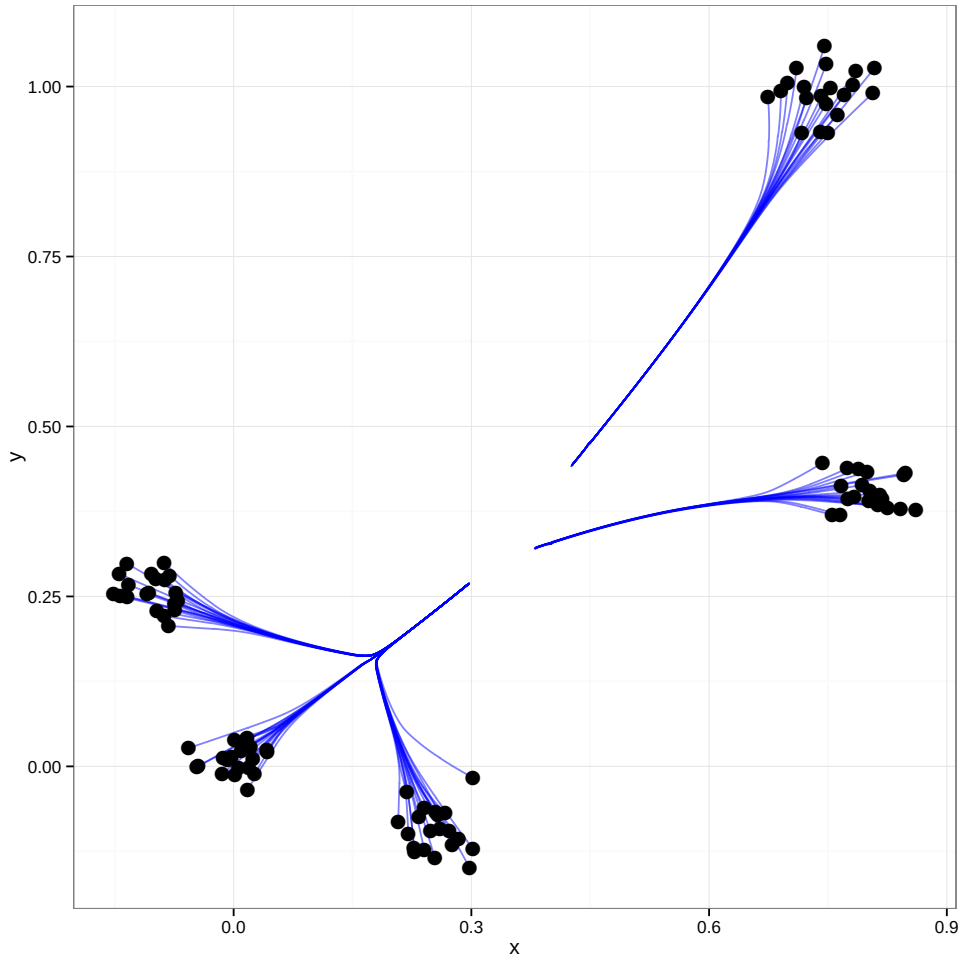
λ

Convex Clustering Solution Path

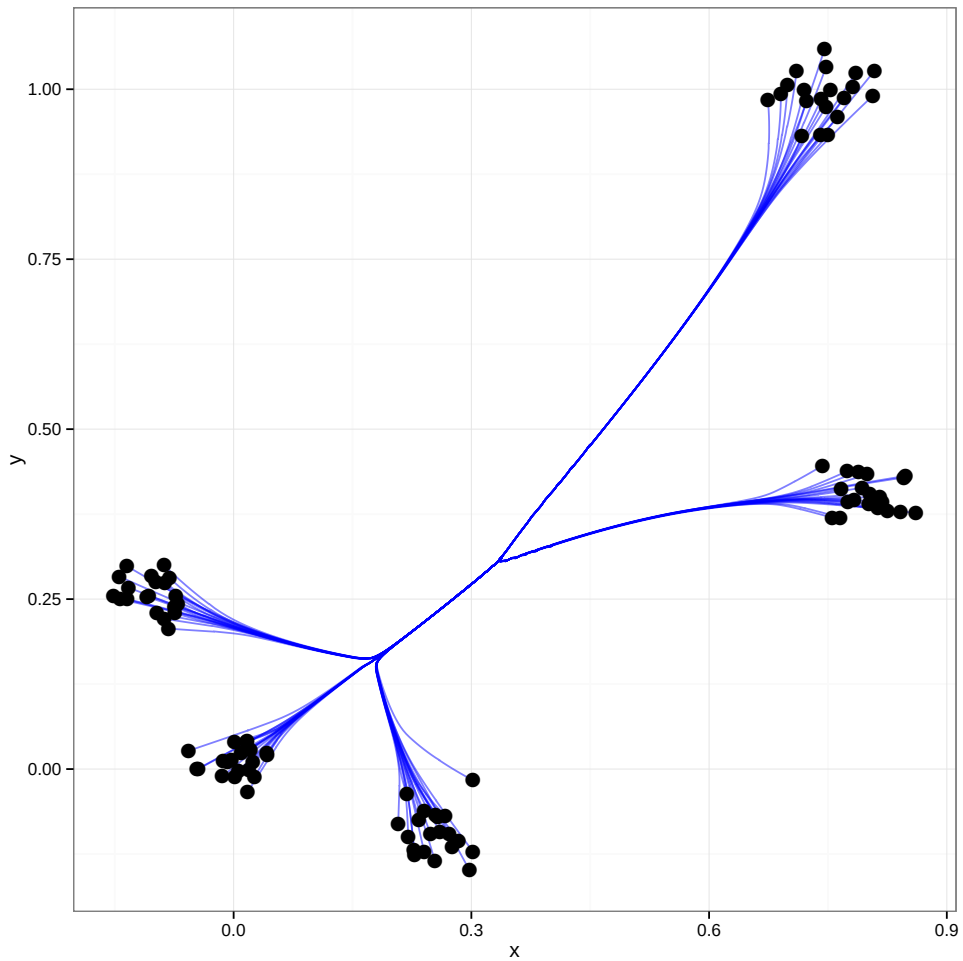


λ

Convex Clustering Solution Path

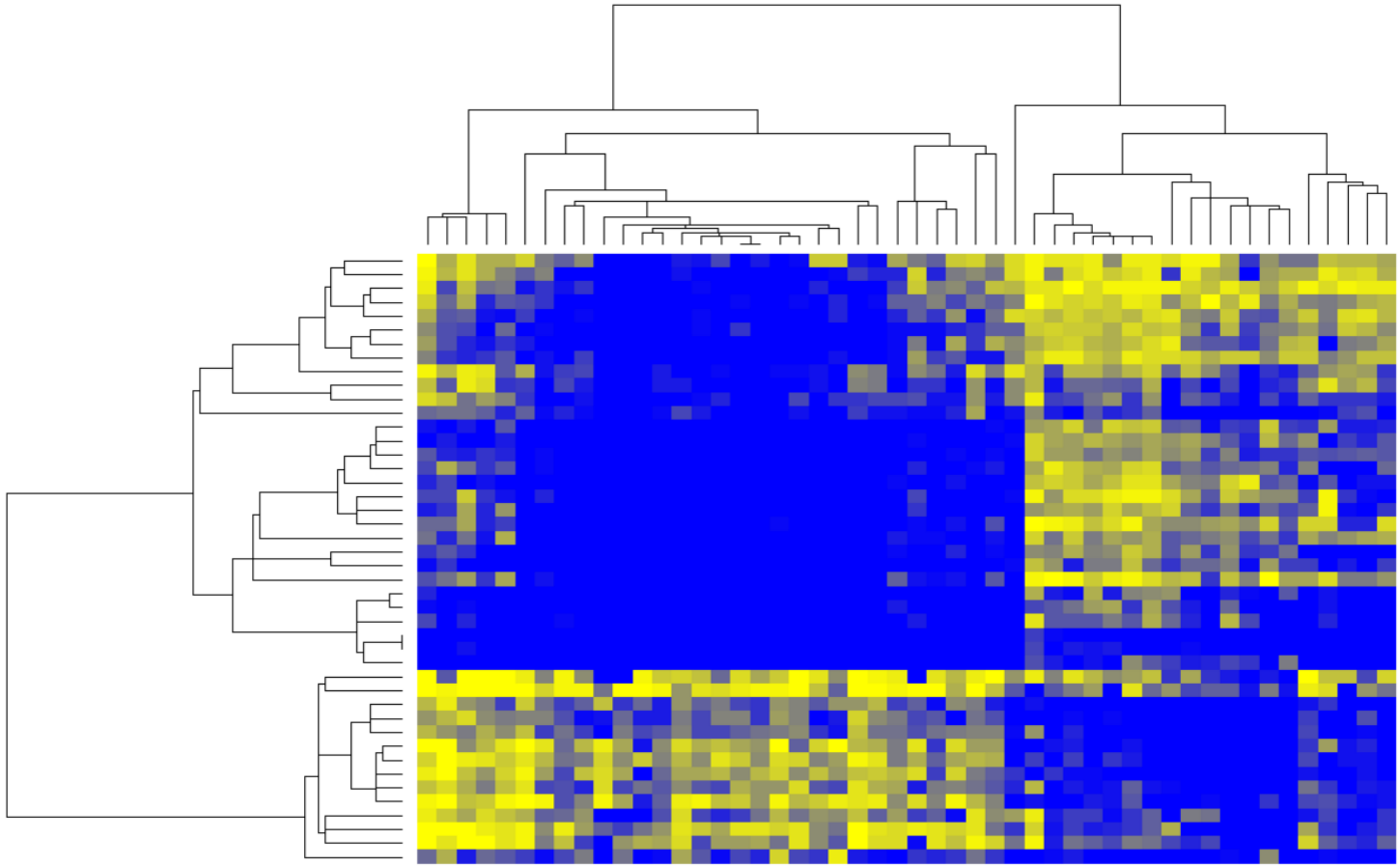


Convex Clustering Solution Path



λ

Convex Biclustering



Convex Biclustering

$$\underset{\mathbf{U}}{\text{minimize}} \quad \frac{1}{2}\|\mathbf{X}-\mathbf{U}\|_F^2 + \lambda\Big(\sum_{i<j} w_{ij}\|\mathbf{U}_{i\cdot}-\mathbf{U}_{j\cdot}\|_2 + \sum_{l<k} \tilde{w}_{lk}\|\mathbf{U}_{\cdot l}-\mathbf{U}_{\cdot k}\|_2\Big)$$

- Checkerboard-like pattern: every data point X_{ij} has its own bicluster centroid U_{ij}.
- Simultaneously fuses row centroids AND column centroids to yield biclusters!

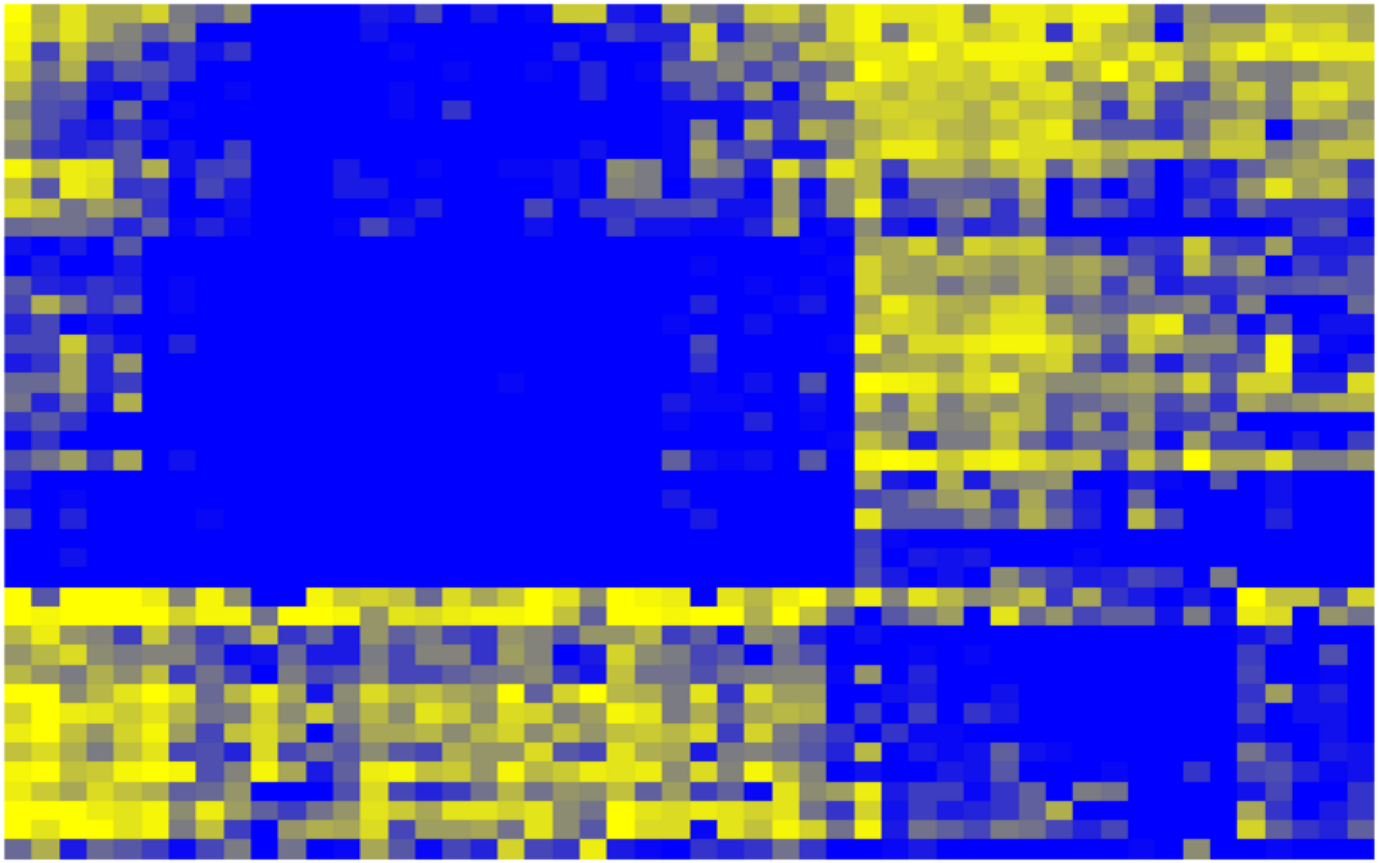
Chi, Allen, and Baraniuk, 2017

Convex Biclustering

$$\underset{\mathbf{U}}{\text{minimize}} \quad \frac{1}{2}\|\mathbf{X}-\mathbf{U}\|_F^2 + \lambda\Bigg(\sum_{i<j} w_{ij}\|\mathbf{U}_{i\cdot}-\mathbf{U}_{j\cdot}\|_2 + \sum_{l<k}\tilde{w}_{lk}\|\mathbf{U}_{\cdot l}-\mathbf{U}_{\cdot k}\|_2\Bigg)$$

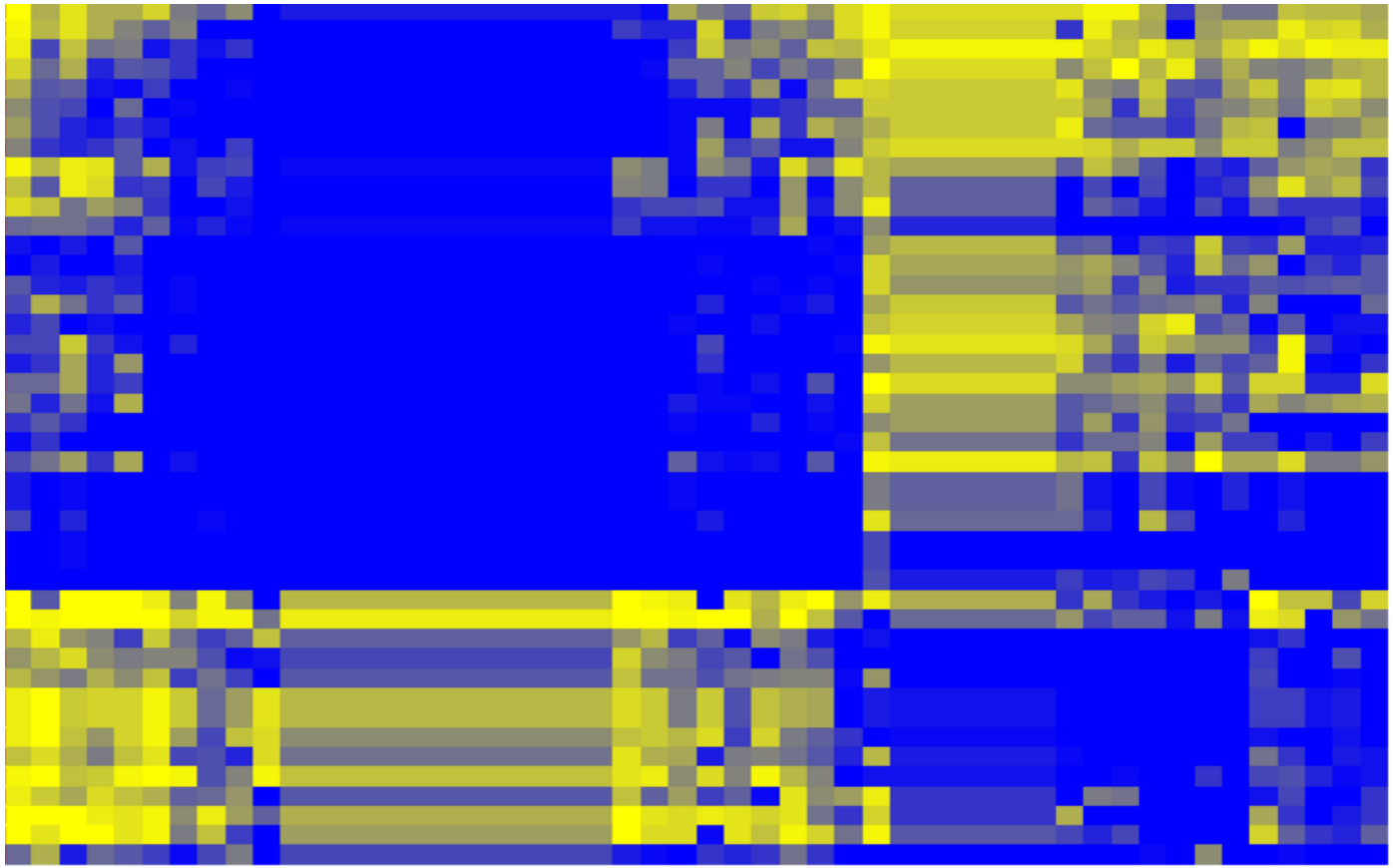
- λ controls BOTH bicluster assignments and # of biclusters.
- Weights similar to convex clustering.
  - Must sum to $1/\sqrt{p}$ and $1/\sqrt{n}$ to ensure the same fusion rate.
- Algorithm: Dystra-like Proximal Algorithm + AMA.
- In R: `cvxbiclustr`.

Convex Biclustering Solution Path



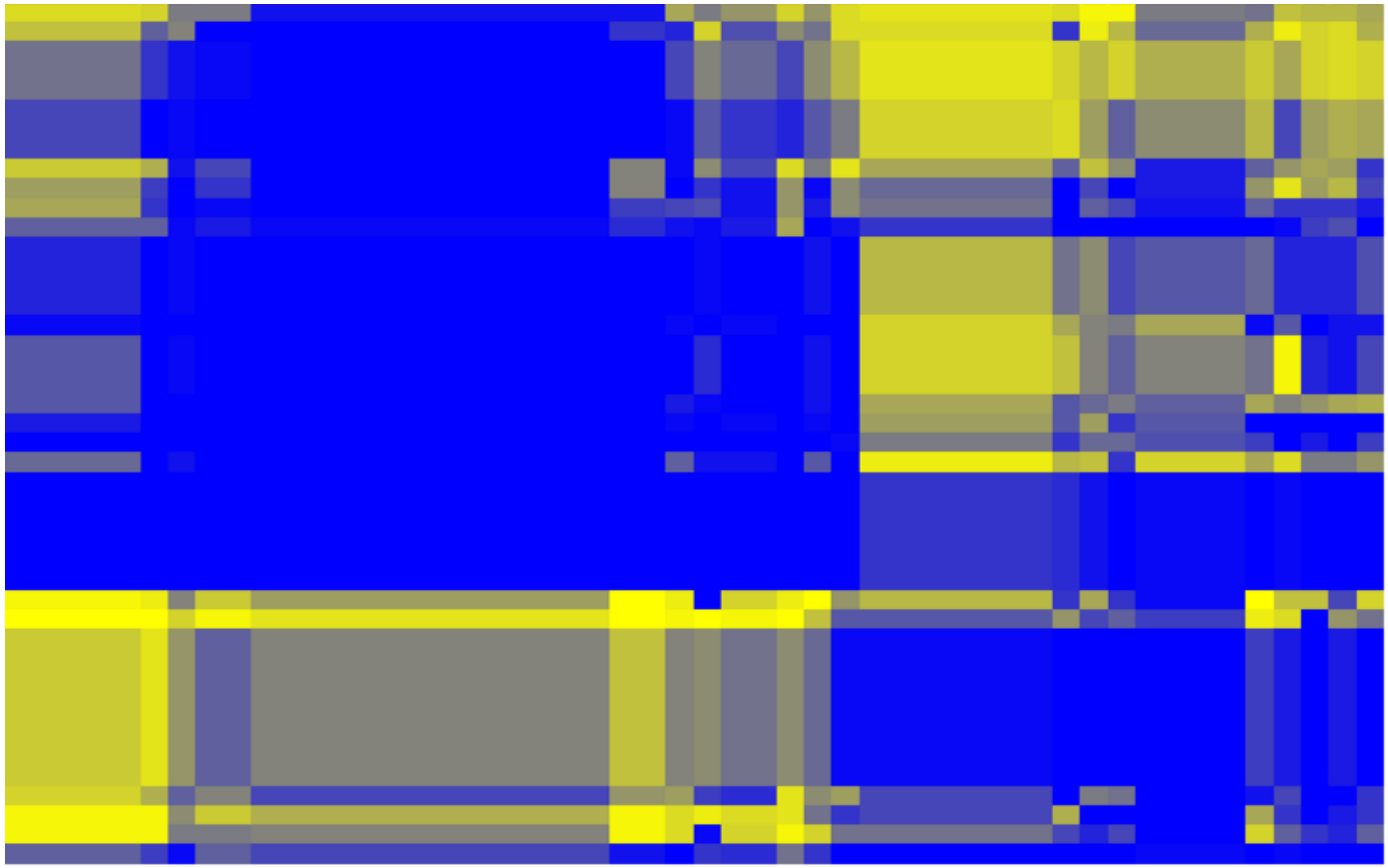
$\lambda = 0$

Convex Biclustering Solution Path



λ

Convex Biclustering Solution Path



λ

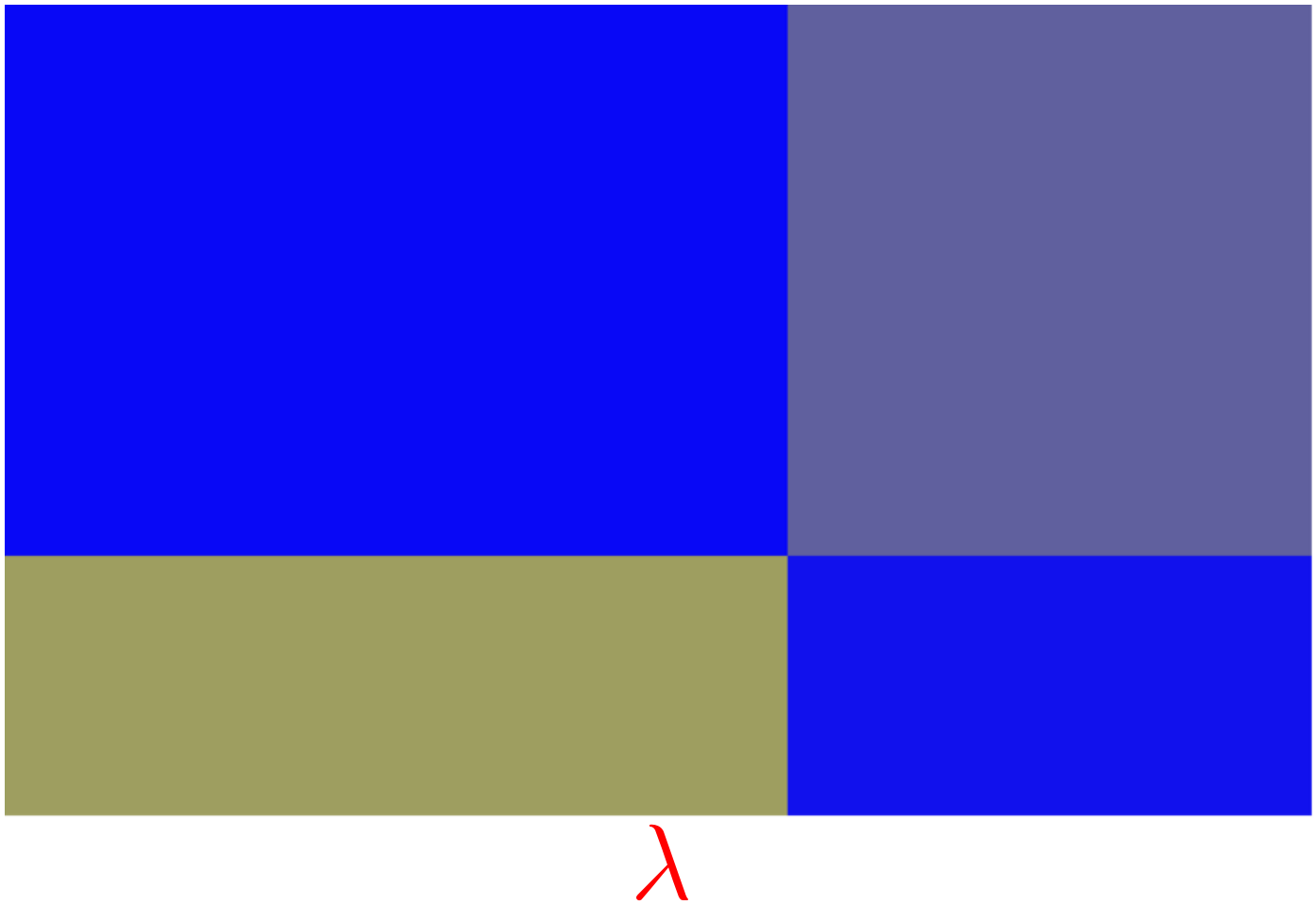
Convex Biclustering Solution Path

λ

Convex Biclustering Solution Path

λ

Convex Biclustering Solution Path



Convex Biclustering Solution Path

λ

Advantages

The Good:

- Global solution!
  - Stable, reproducible results.
- One tuning parameter.
  - λ controls BOTH # of clusters & cluster assignments.
  - Can select in data-driven manner - **Cross Validation**!
- Statistical Consistency.

The Bad:

- Nested family of clustering solutions?
- Slower iterative algorithms to find solution.
- Inference.

Chi, Allen, and Baraniuk, 2017; Tan & Witten 2015; Radchenko & Mukherjee, 2017

Our Objective

Watch your data form clusters & biclusters!

Goal

- Dendograms & Clusterheatmaps.
- Convex clustering & biclustering solution paths.

Our Objective

Our Objective

Watch your data form clusters & biclusters!

Goal

- Dendograms & Clusterheatmaps.
- Convex clustering & biclustering solution paths.

Problems:

- Potential fissions.
  - *Hocking et al. 2011; Tan & Witten 2015*
- Need exact λ where all fusions occur.
  - Existing algorithms solve for one λ at a time.
  - LAR / Path algorithm for Generalized Lasso doesn't work for convex clustering problem.

Computationally way too slow!

Our Objective

Watch your data form clusters & biclusters!

Goal

- Dendograms & Clusterheatmaps.
- Convex clustering & biclustering solution paths.

Our Approach: **Algorithmic Regularization Paths**

- Quickly approximate clustering solution path at a very fine resolution.
  - *Hu, Allen, & Chi, 2017*

Algorithmic Regularization Paths for Clustering

Classical Regularization Paths

Start: Each observation is its own cluster & no regularization.

Do: Increase the regularization level (λ) by a tiny amount.

Do: Solve the optimization problem at λ.

Iterate the AMA updates until convergence.

Stop: All observations fused to one cluster.

Output: Solution at each λ as the Clustering Path.

Algorithmic Regularization Paths for Clustering

Idea

Start: Each observation is its own cluster & no regularization.

Do: Perform one iterate of the AMA.

Do: Increase the regularization level by a tiny amount.

Stop: All observations fused to one cluster.

Output: Iterates as the Algorithmic Clustering Path.

Algorithmic Regularization Paths for Clustering

Idea

Start: Each observation is its own cluster & no regularization.

Do: Perform one iterate of the AMA.

Do: Increase the regularization level by a tiny amount.

Stop: All observations fused to one cluster.

Output: Iterates as the Algorithmic Clustering Path.

CARP

Convex clustering via Algorithmic Regularization Paths

Algorithmic Regularization Paths for Clustering

Idea

Start: Each observation is its own cluster & no regularization.

Do: Perform one iterate of the AMA.

Do: Increase the regularization level by a tiny amount.

Stop: All observations fused to one cluster.

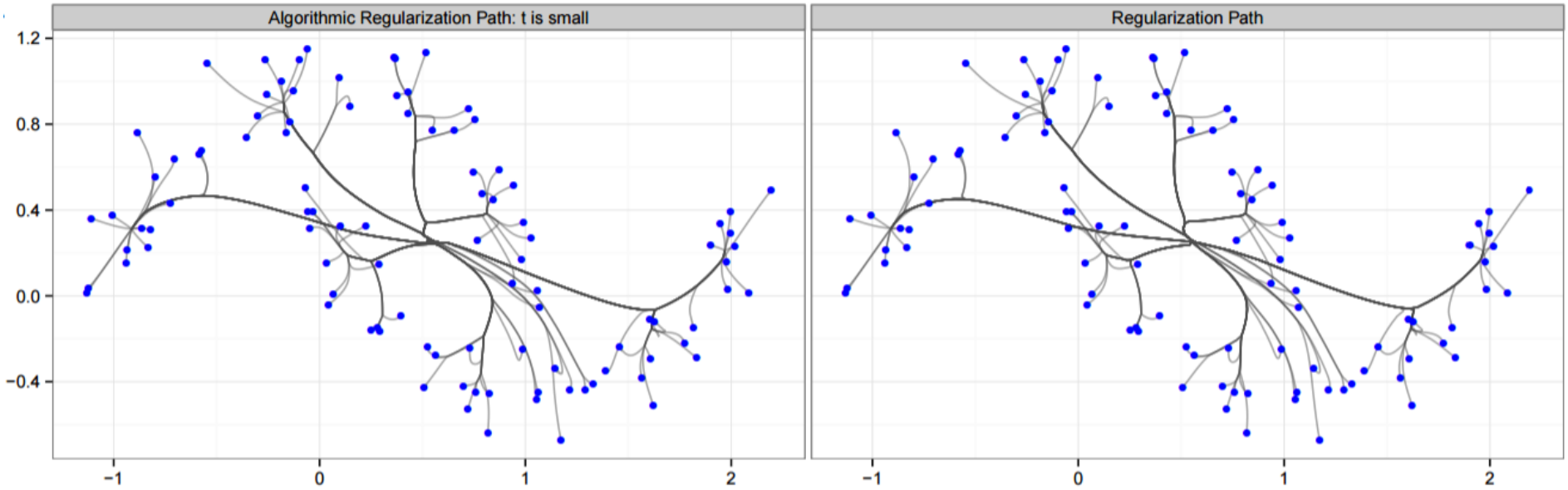
Output: Iterates as the Algorithmic Clustering Path.

CBASS

Convex Biclustering via Algorithmic Regularization with Small Steps

Clustering Path Equivalence

Clustering Path Equivalence for small t:



Very Fast!

Clustering Path Equivalence

Theorem

The algorithmic convex clustering path, $\tilde{\mathbf{U}}_t(k)$, is equivalent to the convex clustering path, $\hat{\mathbf{U}}(\lambda)$, as the step size $t \longrightarrow 1$:

$$d_H(\hat{\mathbf{U}}(\lambda), \tilde{\mathbf{U}}_t(k)) \longrightarrow 0.$$

where $d_H(\hat{\mathbf{U}}(\lambda), \tilde{\mathbf{U}}_t(k))$ is the Hausdorff distance:

$$d_H(\hat{\mathbf{U}}(\lambda), \tilde{\mathbf{U}}_t(k)) = \max\Big\{ \max_k \min_\lambda ||\mathbf{U}(\lambda) - \tilde{\mathbf{U}}_t(k)||_F^2, \max_\lambda \min_k ||\mathbf{U}(\lambda) - \tilde{\mathbf{U}}_t(k)||_F^2 \Big\}.$$

Visualization Algorithm

CARP-VIZ

Convex clustering via Algorithmic Regularization Paths

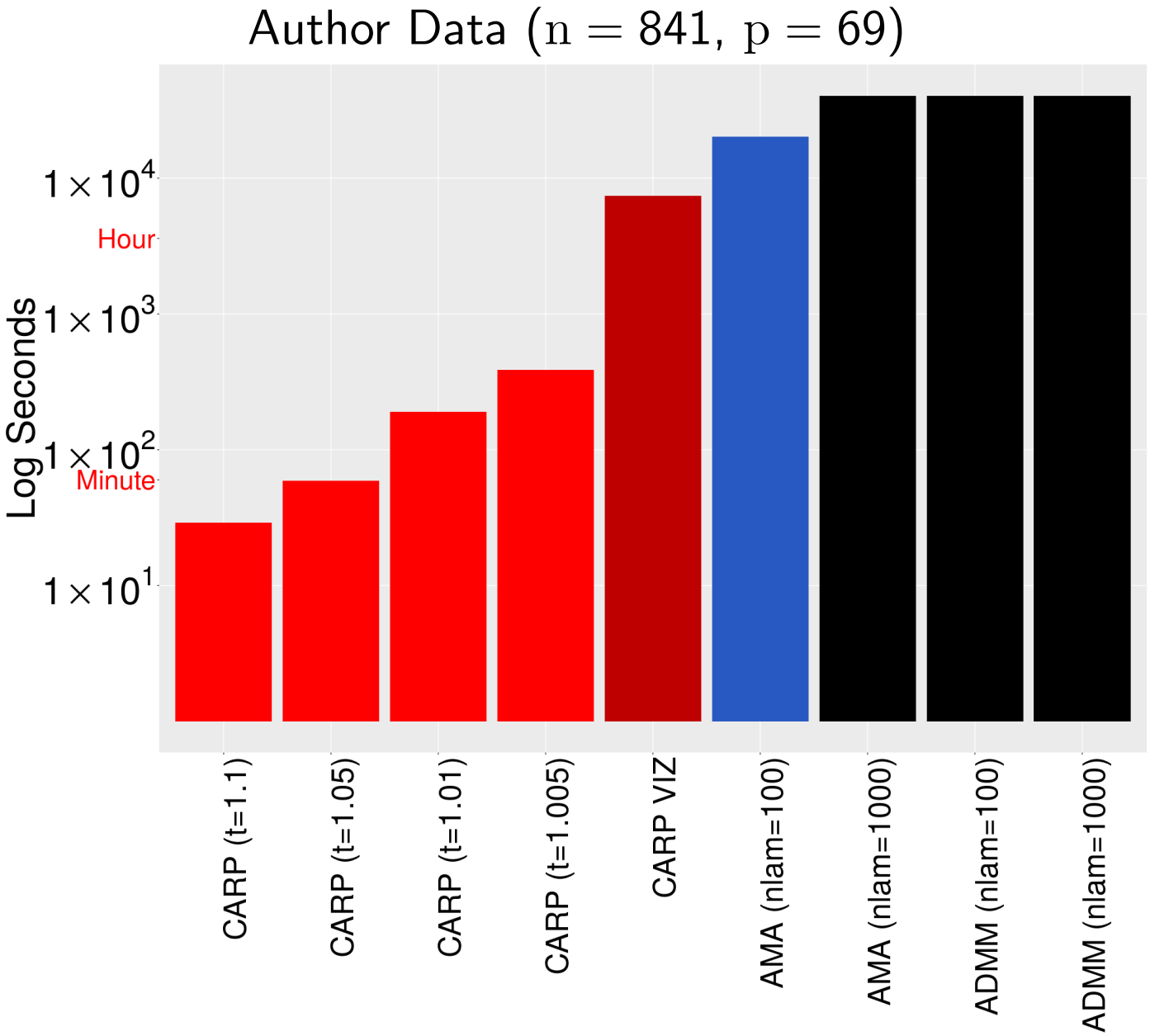
CBASS-VIZ

Convex Biclustering via Algorithmic Regularization with Small Steps

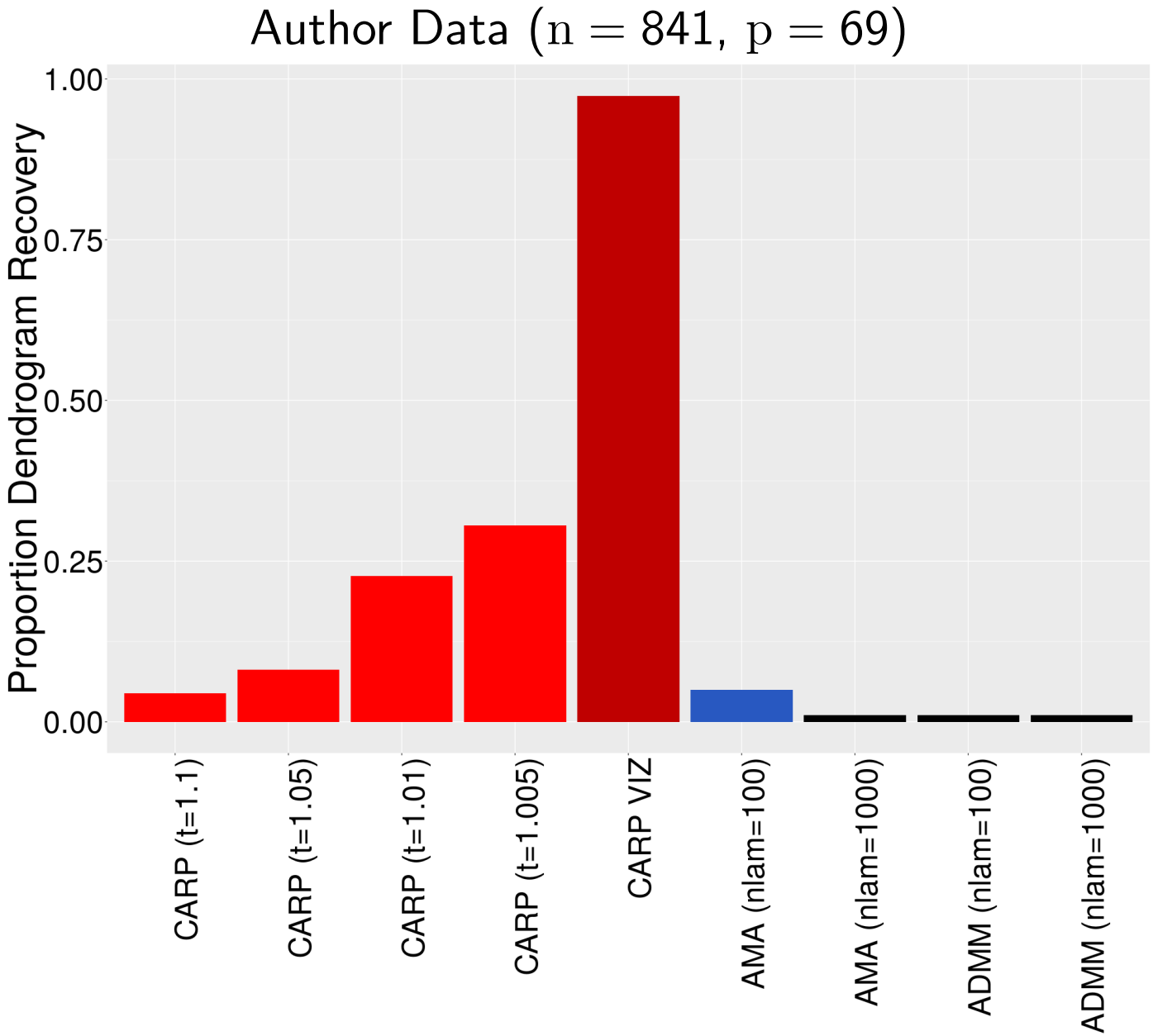
Algorithm to approximate the dendogram:

- Altered proximal operator to prevent fissions.
- Adaptively increase the amount of regularization to find each fusion.

Timing Comparisons



Timing Comparisons



Text Mining Example

US Presidential Speeches Data Set:

Text Mining Example

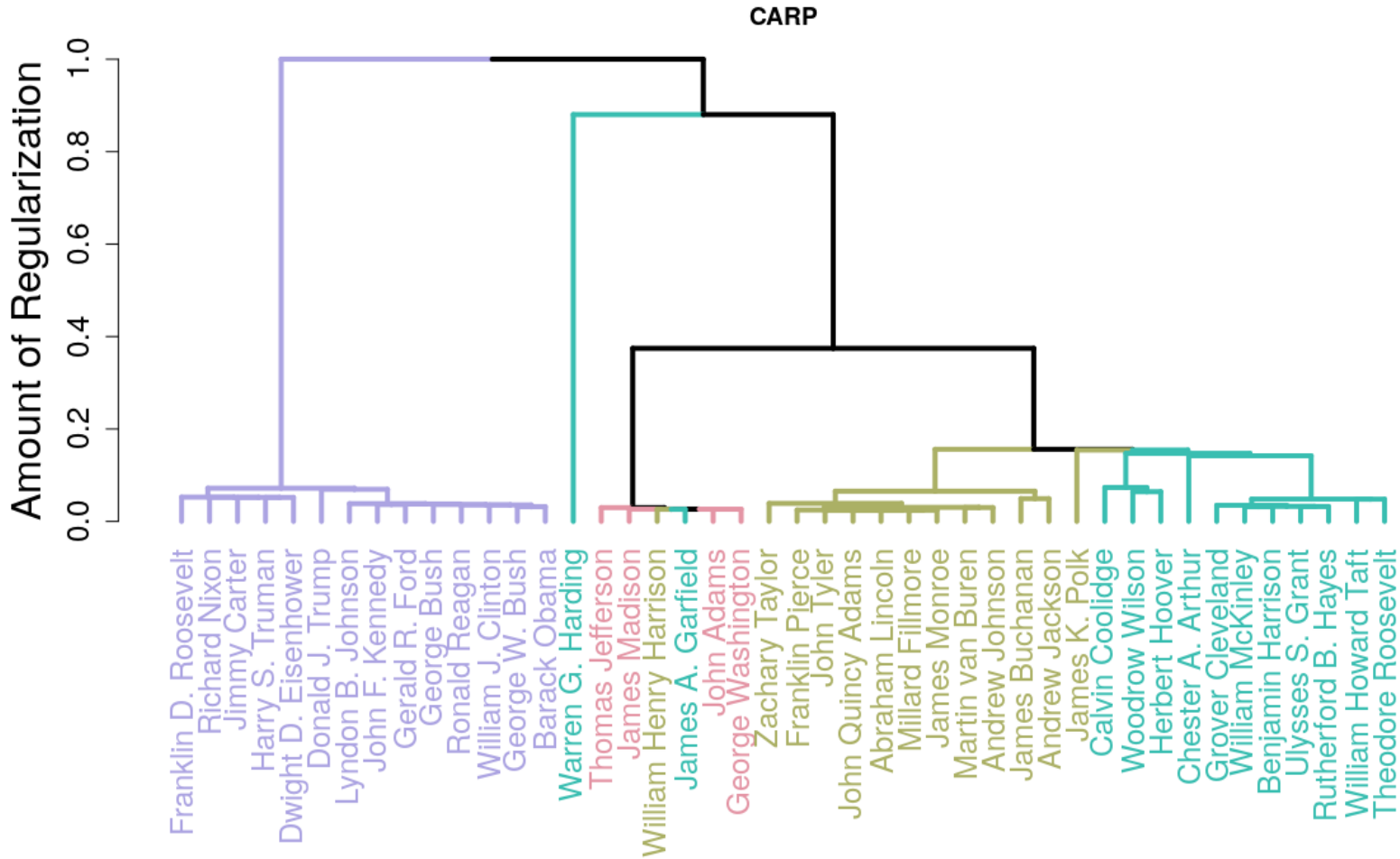
US Presidential Speeches Data Set:

- Text of Inauguration and State of the Union speeches scraped from web: `http://www.presidency.ucsb.edu`.

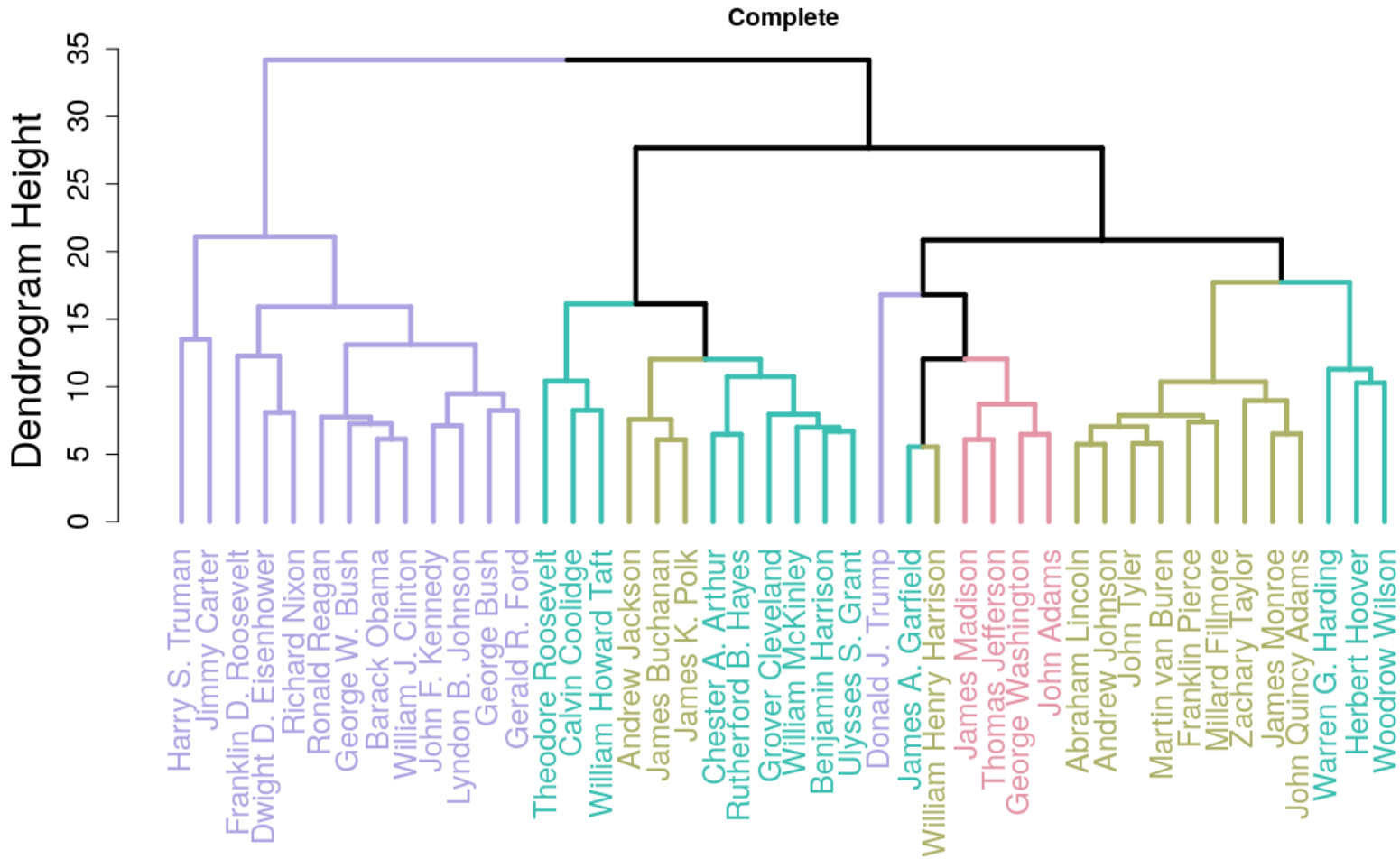
- Processed into a document-word count matrix using the `tm`, ‘‘`Text Mining`’’, `R` package.
  - Lower case; remove white space, stop words, symbols; stem words (remove prefixes and suffixes), etc.

- Filter to top 75 most variable words.
- Log-transform & center and scale whole matrix.

CARP Results - Presidents Data



CARP Results - Presidents Data



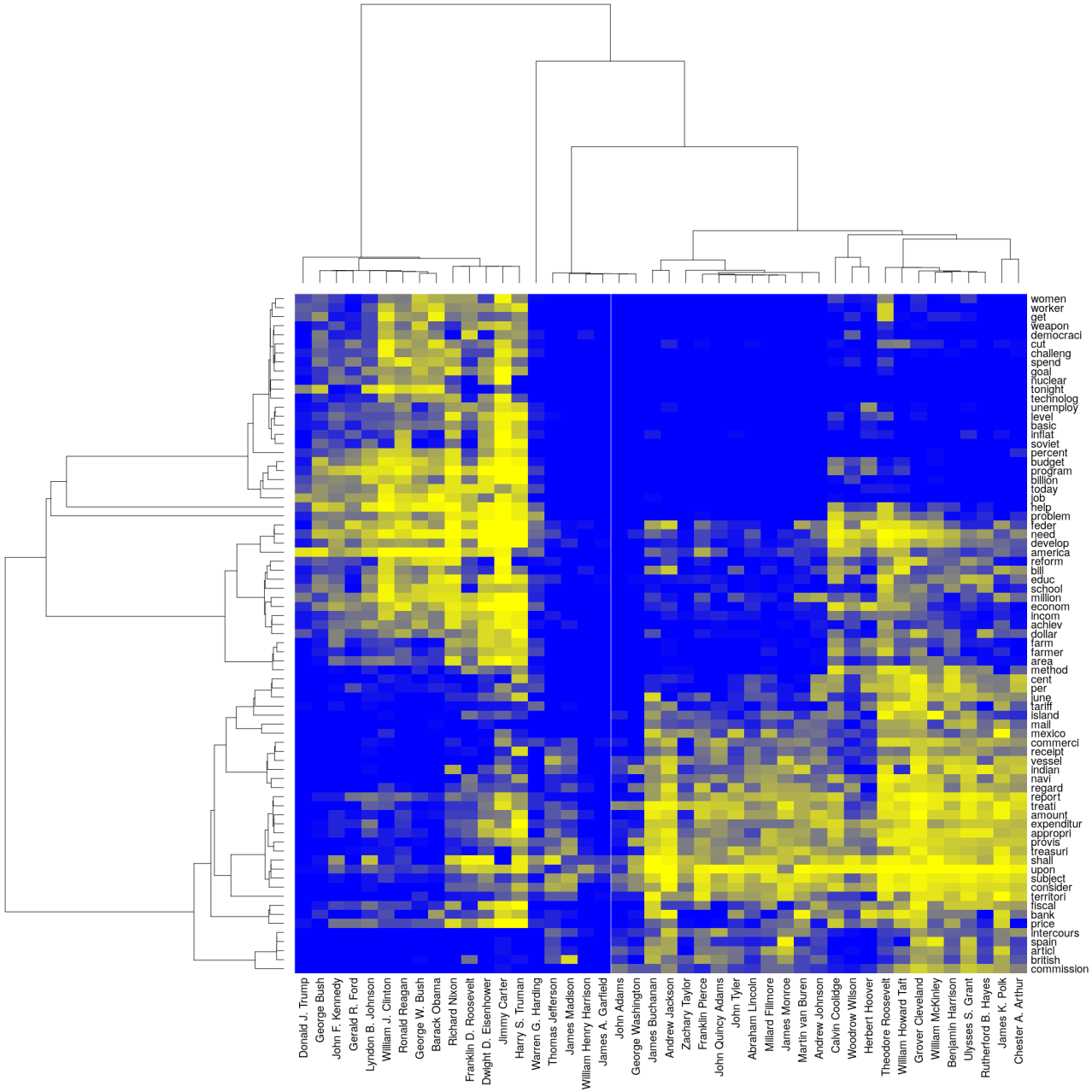
Hierarchical Clustering - Complete Linkage

CARP Results - Presidents Data

CARP Results - Presidents Data

CBASS Results - Presidents Data

CBASS Results - Presidents Data



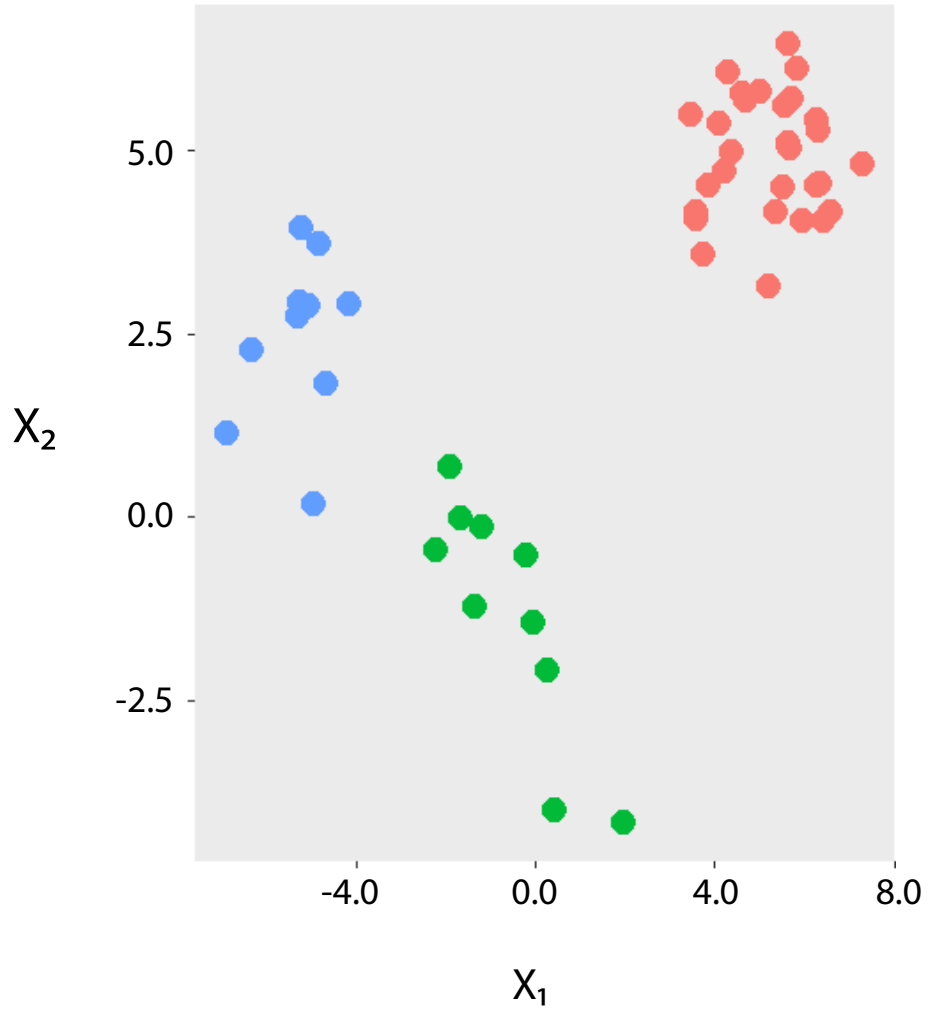
Software

clustRviz

Coming soon!

Inference for Clustering

- Are there true clusters in my data?
- How many clusters?



Heller and Ghahramani (2005); Liu et al. (2008); Kimes et al. (2017); Huang et al. (2015); Hyun et al. (2016)

Our Work

We derive exact null distributions for the following tests:

- One-Sample Test. Conditional on the estimated convex clustering solution, test cluster $\mathrm{g_k}$'s mean:

$$\mathrm{H_0}: \quad \boldsymbol{\mu}_\mathrm{k} = \boldsymbol{\mu}_0 \quad \text{vs.} \quad \mathrm{H_A}: \boldsymbol{\mu}_\mathrm{k} \neq \boldsymbol{\mu}_0 \;\Big|\; \text{Convex Clustering Solution}$$

$\Longrightarrow$ Confidence regions for cluster means.

- Two-Sample Test. Conditional on the estimated convex clustering solution, test whether cluster $\mathrm{g_k}$ and $\mathrm{g_j}$ have equal means:

$$\mathrm{H_0}: \quad \boldsymbol{\mu}_\mathrm{k} = \boldsymbol{\mu}_\mathrm{j} \quad \text{vs.} \quad \mathrm{H_A}: \boldsymbol{\mu}_\mathrm{k} \neq \boldsymbol{\mu}_\mathrm{j} \;\Big|\; \text{Convex Clustering Solution}$$

$\Longrightarrow$ Test whether two clusters are truly separate.

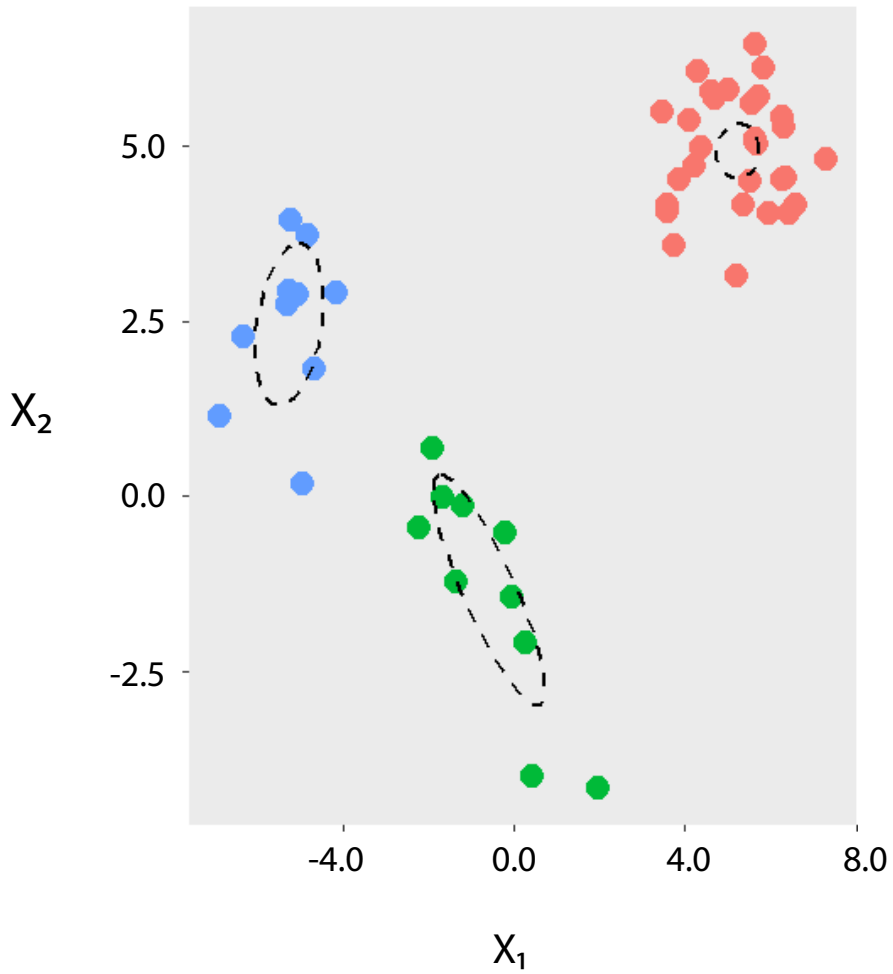
Our Work

- Derive novel representations of one-sample & two-sample Hotelling's T^2 statistic in terms of principal angles.
- Derive novel data decompositions as a function of principal angles.
- Selective inference framework to derive exact null distributions.
  - Post Selection Inference.

Skipping the math ...

Inference for Convex Clustering: Example

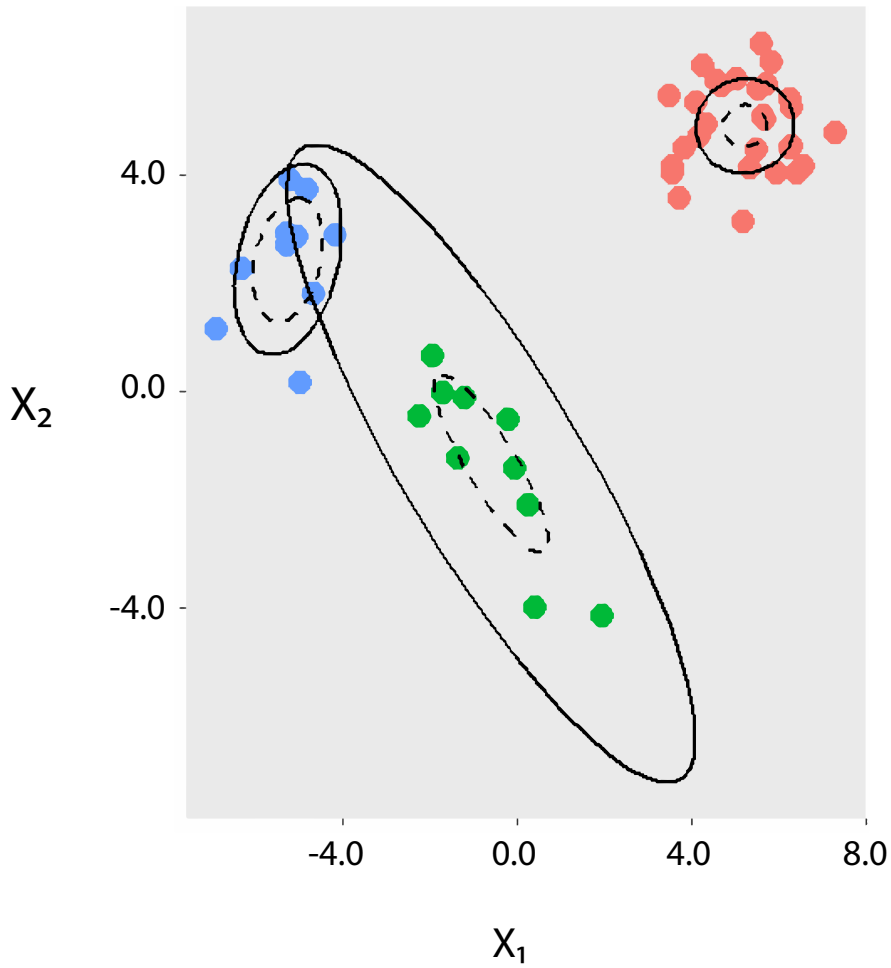
- Confidence ellipsoids for cluster means.
  - Naive: dashed lines.
- Two sample test for equality of cluster green and blue means.
  - Naive:
    p-value $=$ 1.683517e-07



Inference for Convex Clustering: Example

- Confidence ellipsoids for cluster means.
  - Naive: dashed lines.
  - Ours: solid lines.

- Two sample test for equality of cluster green and blue means.
  - Naive:
    p-value $=$ 1.683517e-07
  - Ours:
    p-value $=$ 0.1598832



Summary

Summary

1. Convex Clustering & Biclustering have many advantages!
2. Developed a fast algorithm to compute cluster solution path & approximate cluster dendogram.
   - Novel approach: Algorithmic Regularization Paths.
3. Developed interactive & dynamic visualizations for clustering and biclustering.
4. Developed valid inference procedures for convex clustering.
   - Novel representations of data in terms of principal angles.

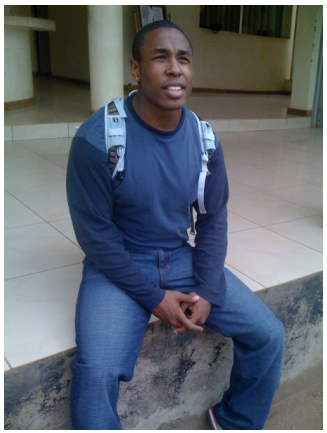
clustRviz

Coming soon!

Acknowledgments & References

John Nagorski,
PhD, Statistics, Rice University

Frederick Campbell,
PhD, Statistics, Rice University



Acknowledgments & References

E. C. Chi, G. I. Allen, and R. Baraniuk, "Convex Biclustering", **73**:1, 10-19, *Biometrics*, 2017.

Y. Hu, E. C. Chi, and G. I. Allen, "ADMM Algorithmic Regularization Paths for Sparse Statistical Machine Learning", In *Splitting Methods in Communication and Imaging, Science and Engineering*, R. Glowinski, W. Yin, and S. Osher (eds), 2017.

J. Nagorski, M. Weylandt, and G. I. Allen, "Dynamic Visualization and Fast Computation of the Solution Path for Convex Clustering", *Preprint*, 2018.

F. Campbell and G. I. Allen, "Inference for Multivariate Means in Adaptive Data Analysis", *Working Paper*, 2018.

Thank You!