Unboxing Cluster Heatmaps

Sophie Engle   **Sean Whalen**
Alark Joshi   **Katherine Pollard**

vgl.cs.usfca.edu   docpollard.org
Visualization and Graphics Lab   Pollard Group
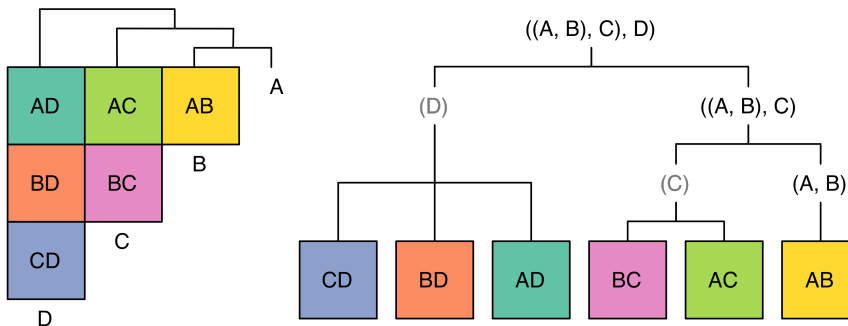University of San Francisco   Gladstone Institutes, UCSF

Contributions

- Practitioner Survey

- 45 practitioners

- Heatmap & dendrograms

- Symmetric & asymmetric

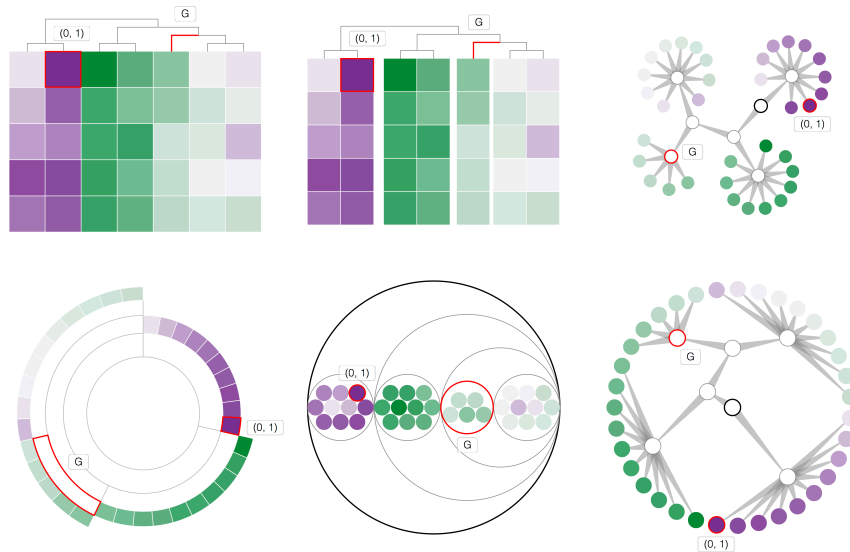- 100 to 250,000 cells

- R & Cytoscape

Unboxing Cluster Heatmaps by Engle, Whalen, Joshi, and Pollard
6th Symposium on Biological Data Visualization (BioVis), October 23, 2016

Visualization and Graphics Lab, Department of Computer Science, University of San Francisco
Pollard Group, Gladstone Institutes, University of California San Francisco

Contributions

- Practitioner Survey

- Unboxing Approach

Unboxing Cluster Heatmaps by Engle, Whalen, Joshi, and Pollard
6th Symposium on Biological Data Visualization (BioVis), October 23, 2016

Visualization and Graphics Lab, Department of Computer Science, University of San Francisco
Pollard Group, Gladstone Institutes, University of California San Francisco

Contributions

- Practitioner Survey

- Unboxing Approach

- Pair Analytics

Unboxing Cluster Heatmaps by Engle, Whalen, Joshi, and Pollard
6th Symposium on Biological Data Visualization (BioVis), October 23, 2016

Visualization and Graphics Lab, Department of Computer Science, University of San Francisco
Pollard Group, Gladstone Institutes, University of California San Francisco

Contributions

- Practitioner Survey

- Unboxing Approach

- Pair Analytics

- Practitioner Interviews

Unboxing Cluster Heatmaps by Engle, Whalen, Joshi, and Pollard
6th Symposium on Biological Data Visualization (BioVis), October 23, 2016

Visualization and Graphics Lab, Department of Computer Science, University of San Francisco
Pollard Group, Gladstone Institutes, University of California San Francisco

Contributions

- Practitioner Survey

- Unboxing Approach

- Pair Analytics

- Practitioner Interviews

- **Large-Scale User Study**
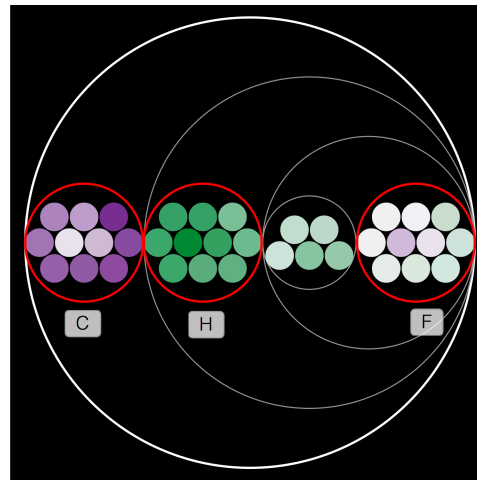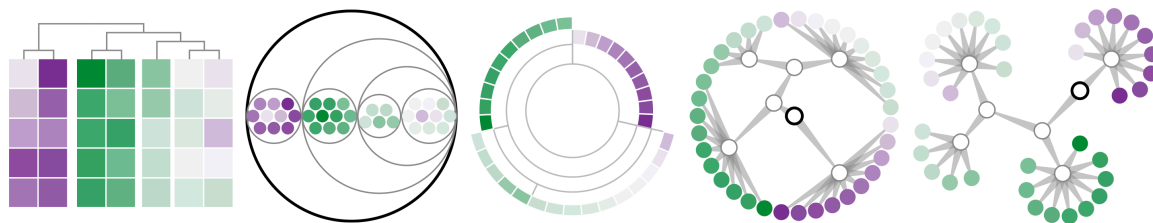
Which two elements are more closely clustered?

Contributions

- Practitioner Survey **45**

- Unboxing Approach

- Pair Analytics

- Practitioner Interviews **5**

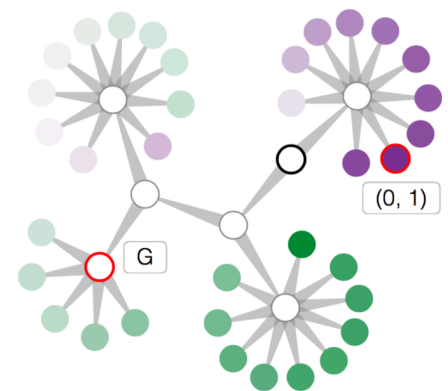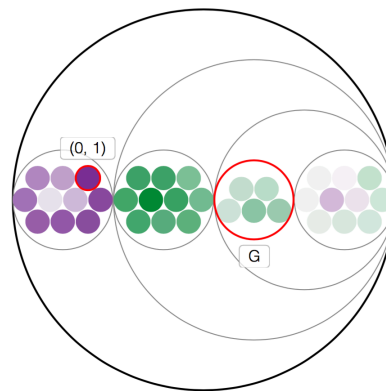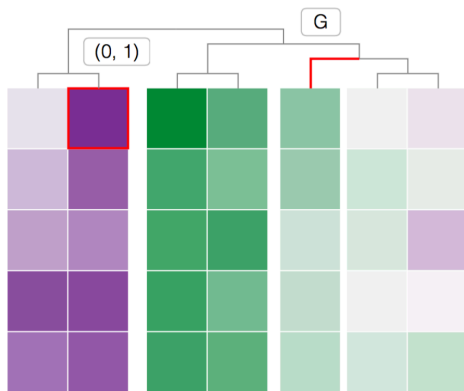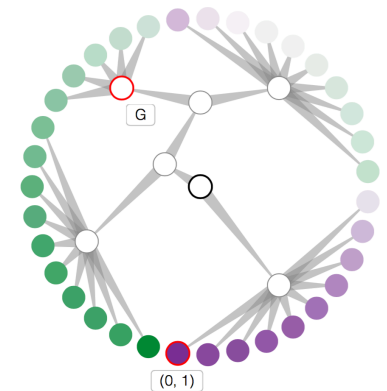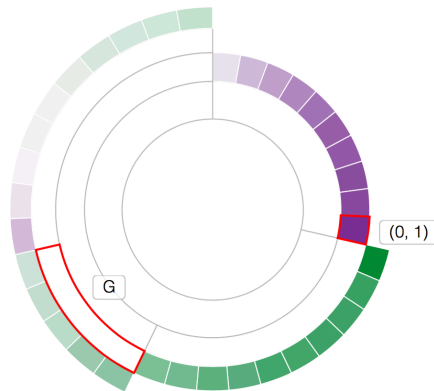- Large-Scale User Study **200**

Contributions

- Practitioner Survey 45

- Unboxing Approach

- Pair Analytics

- Practitioner Interviews 5

- Large-Scale User Study 200

Unboxing Cluster Heatmaps by Engle, Whalen, Joshi, and Pollard
6th Symposium on Biological Data Visualization (BioVis), October 23, 2016
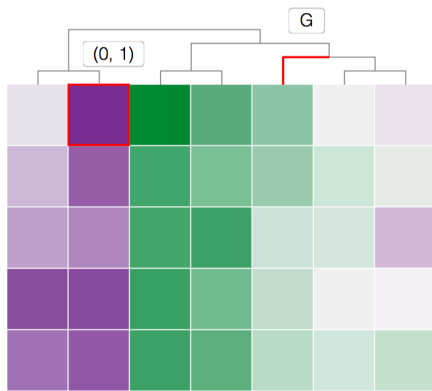
Visualization and Graphics Lab, Department of Computer Science, University of San Francisco
Pollard Group, Gladstone Institutes, University of California San Francisco

Unboxing Cluster Heatmaps

github.com/usfvgl/unboxing-cluster-heatmaps
git.io/vw0t3

Visualization and Graphics Lab, Department of Computer Science, University of San Francisco
Pollard Group, Gladstone Institutes, University of California San Francisco

Left: Cluster Heatmap/Gapmap, **Middle:** Sunburst/Circle Packing, **Right:** Radial Dendrogram/Force-Directed Tree
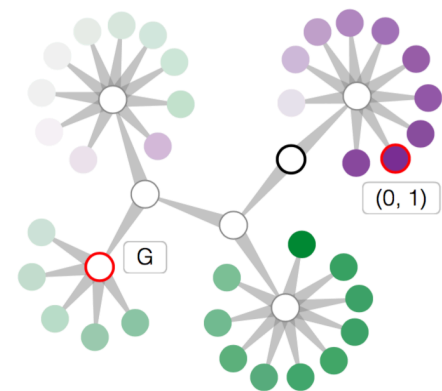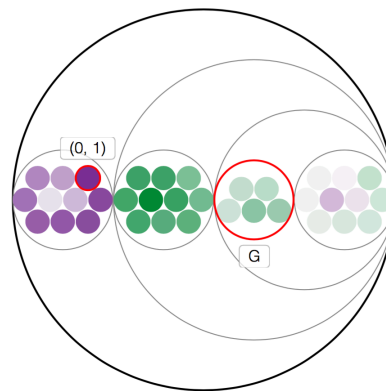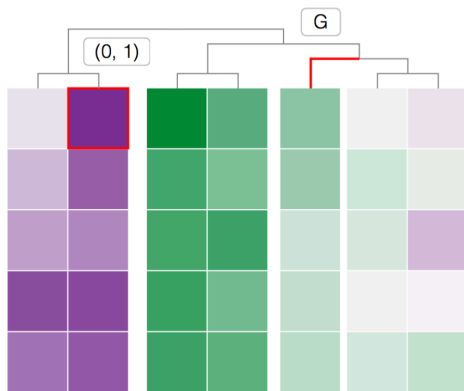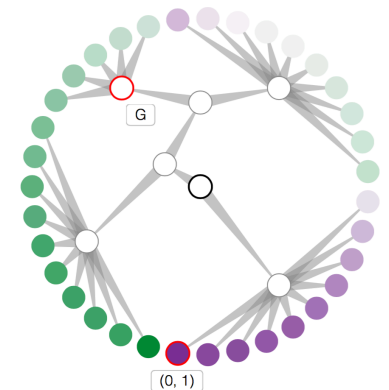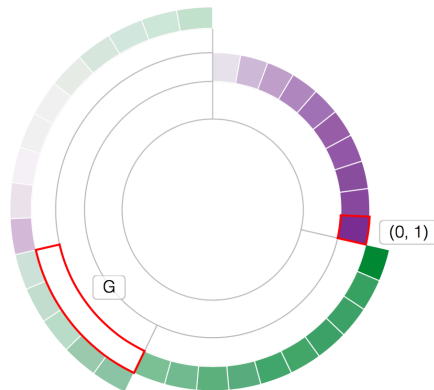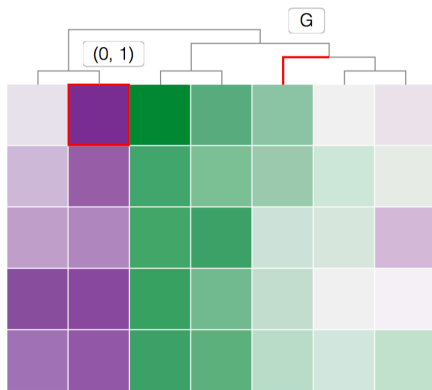
Left: Cluster Heatmap/Gapmap, Middle: Sunburst/Circle Packing, Right: Radial Dendrogram/Force-Directed Tree

mTurk Study Design

Which two of the highlighted elements are more closely clustered?
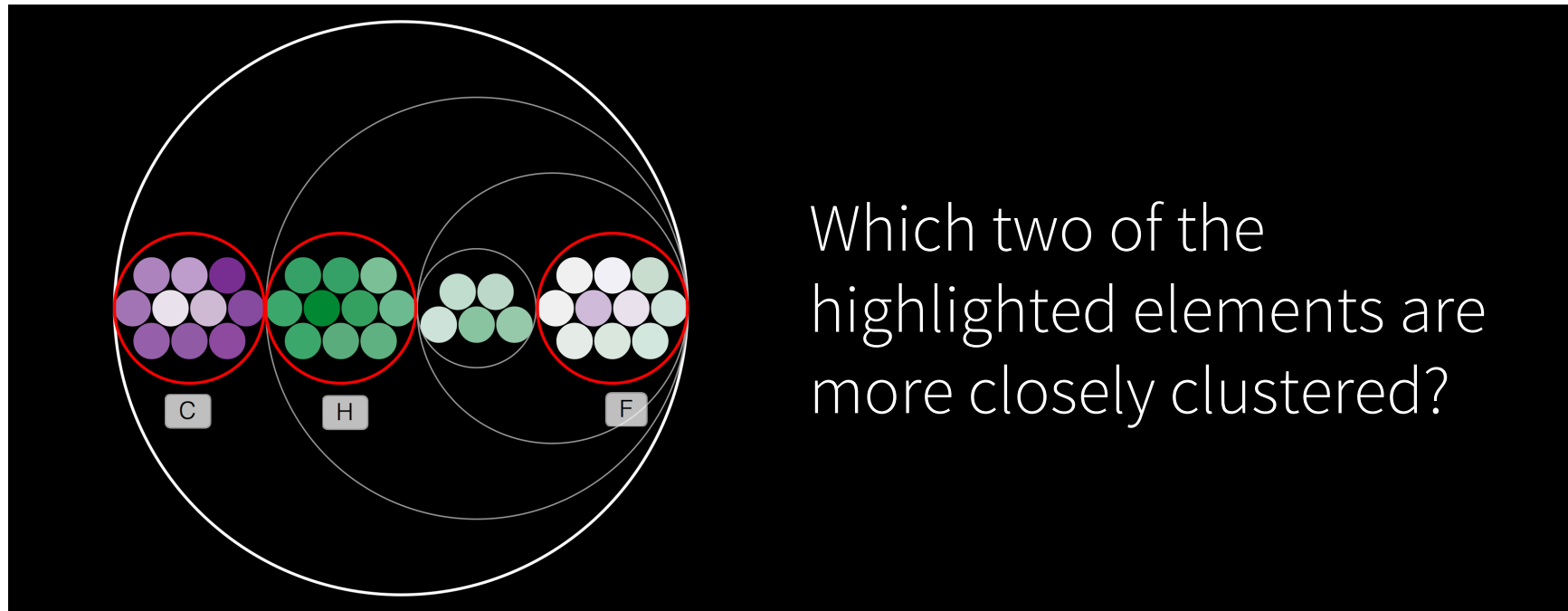
5 by 7 circle packing

mTurk Study Design

Which two of the highlighted elements are more closely clustered?

5 by 7 radial dendrogram

Unboxing Cluster Heatmaps by Engle, Whalen, Joshi, and Pollard
6th Symposium on Biological Data Visualization (BioVis), October 23, 2016

Visualization and Graphics Lab, Department of Computer Science, University of San Francisco
Pollard Group, Gladstone Institutes, University of California San Francisco

mTurk Study Design

(1, 7)

(3, 9)

(6, 8)

Which two of the highlighted elements are more closely clustered?

8 by 16 cluster heatmap

mTurk Study Design

Which two of the highlighted elements are more closely clustered?
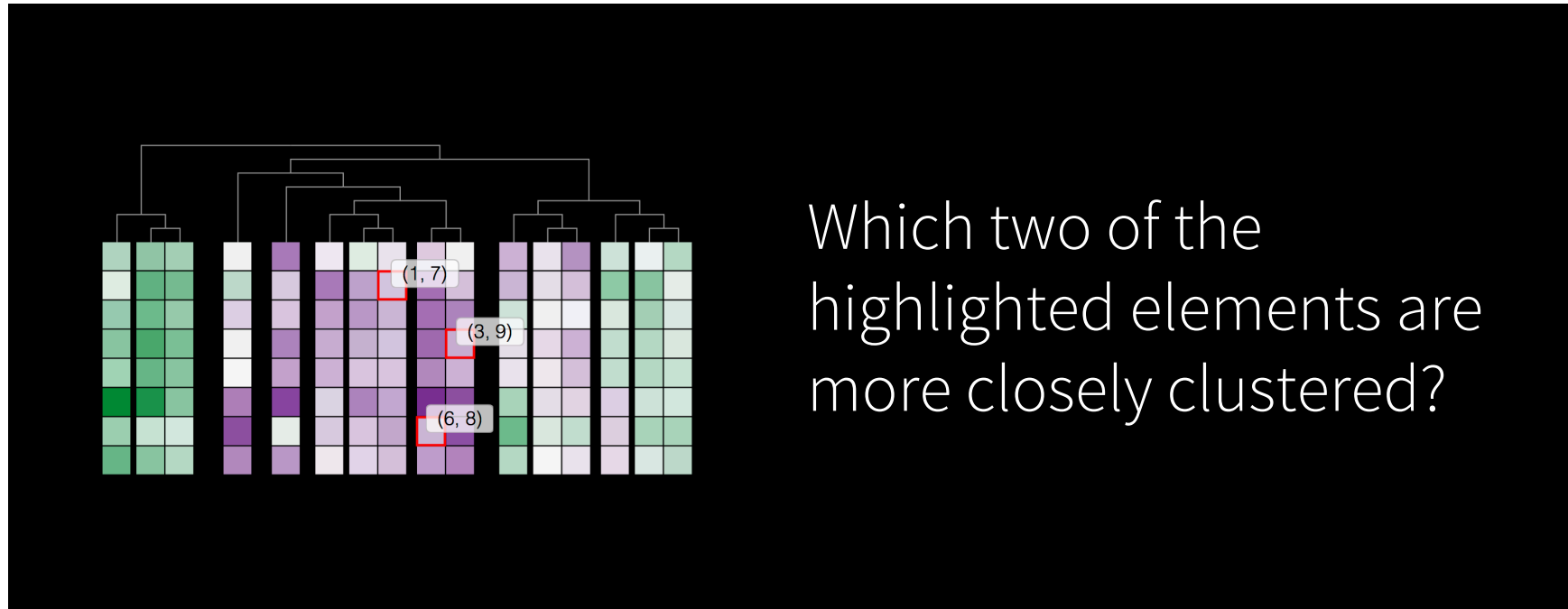
8 by 16 gapmap

mTurk Study Results
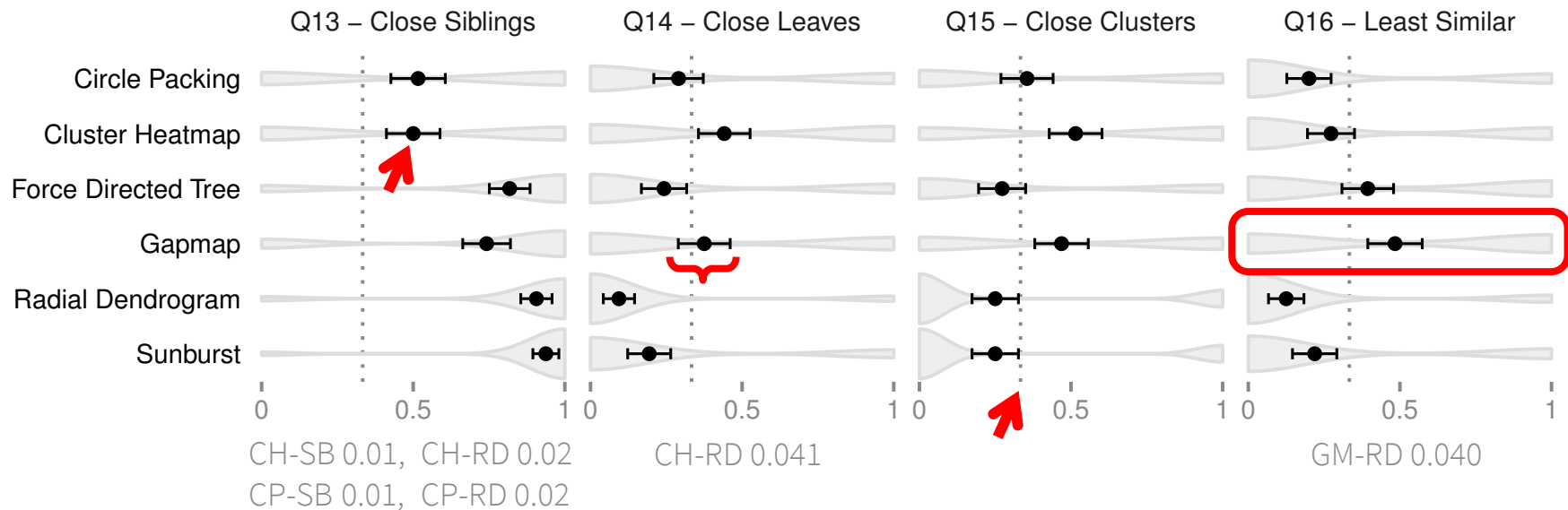
Average Score Violin Plot. Circle: Average Score, Bar: Standard Error, Dotted Line: Random Performance (w/o "Unsure")

Unboxing Cluster Heatmaps by Engle, Whalen, Joshi, and Pollard
6th Symposium on Biological Data Visualization (BioVis), October 23, 2016

Visualization and Graphics Lab, Department of Computer Science, University of San Francisco
Pollard Group, Gladstone Institutes, University of California San Francisco

mTurk Study Results

Average Score Violin Plot. Circle: Average Score, Bar: Standard Error, Dotted Line: Random Performance (w/o "Unsure")
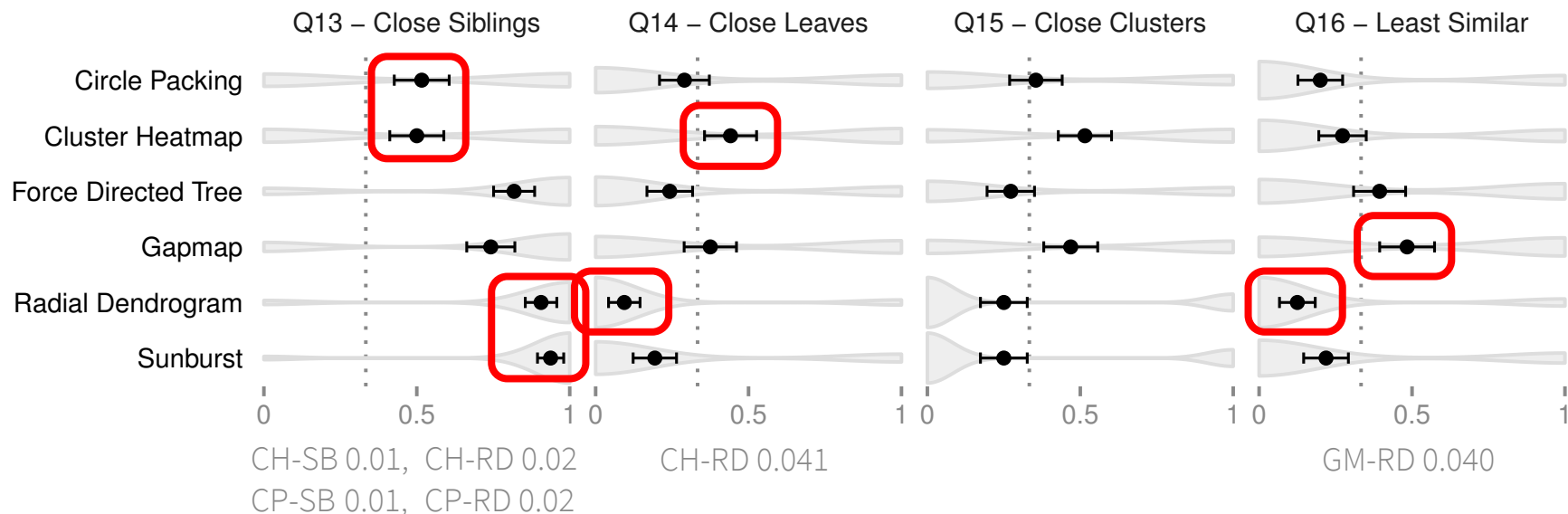
mTurk Study Results

Average Score Violin Plot. Circle: Average Score, Bar: Standard Error, Dotted Line: Random Performance (w/o "Unsure")
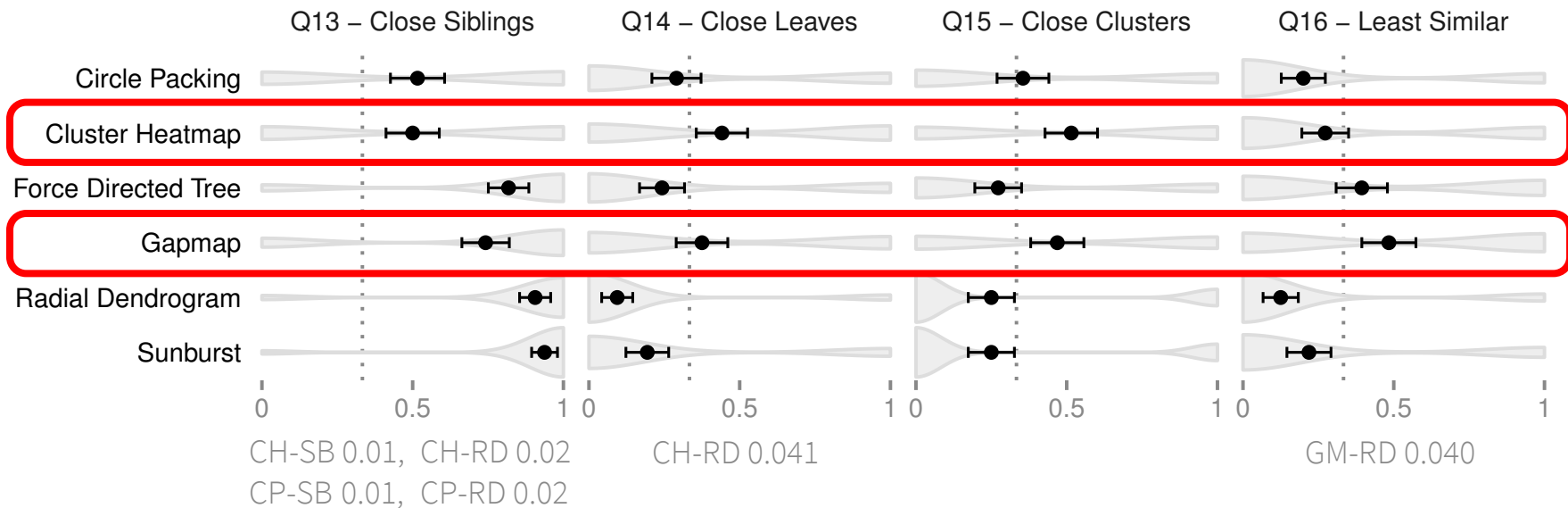
mTurk Study Results

Q17 – Visual Clusters
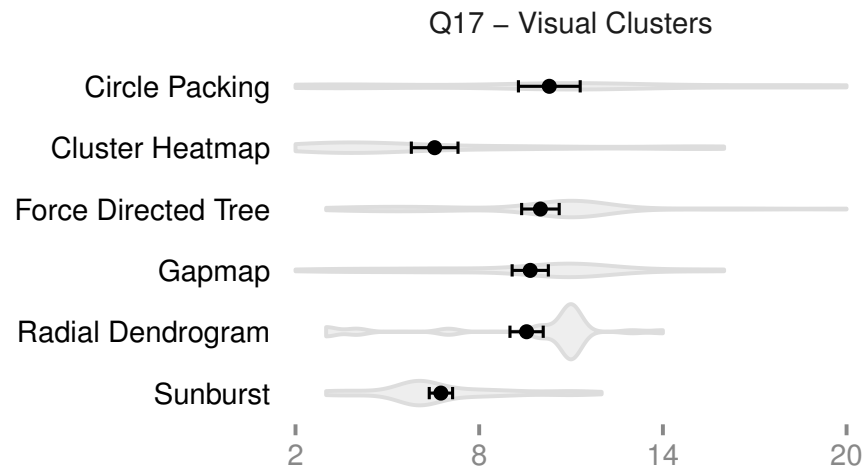
Average Clusters Violin Plot. Circle: Average Estimate, Bar: Standard Error

Conclusions

- Gapmaps are a good alternative for cluster heatmaps
  - Preferred by interviewed practitioners
  - Performed well in mTurk user study

- Involving practitioners at multiple stages critical
  - Caught places were assumptions were incorrect

- Symmetric matrices need (much) more study ★

Unboxing Cluster Heatmaps by Engle, Whalen, Joshi, and Pollard
6th Symposium on Biological Data Visualization (BioVis), October 23, 2016

Visualization and Graphics Lab, Department of Computer Science, University of San Francisco
Pollard Group, Gladstone Institutes, University of California San Francisco

Unboxing Cluster Heatmaps

github.com/usfvgl/unboxing-cluster-heatmaps
git.io/vw0t3

Visualization and Graphics Lab, Department of Computer Science, University of San Francisco
Pollard Group, Gladstone Institutes, University of California San Francisco

Unboxing Approach for a Symmetric Matrix

Type	Task	#	Question Text	Nodes	Type	mean	χ^2	df	p-value
Training	1	1	Is the highlighted cluster mostly positive or mostly negative?	Clusters	N/A				
Training	1	2	What is the height of the tree?	N/A	N/A				
Timed	1	3	Which of the highlighted elements has the highest value?	Leaves	Score	0.380	4.578	5	4.695 E-01
Timed	1	4	Is the highlighted cluster mostly positive or mostly negative?	Clusters	Score	0.661	14.650	5	1.197 E-02 ∗
Timed	1	5	What is the height of the tree?	N/A	Error	3.083	23.278	5	2.987 E-04 ∗∗∗
Training	2	6	Which of the highlighted elements is furthest away from the root?	Leaves	N/A				
Training	2	7	Which of the highlighted elements are siblings?	Leaves	N/A				
Timed	2	8	Which of the highlighted elements is furthest away from the root?	Leaves	Score	0.605	63.540	5	2.250 E-12 ∗∗∗
Timed	2	9	Which of the highlighted elements is furthest away from the root?	Clusters	Score	0.732	14.705	5	1.170 E-02 ∗
Timed	2	10	Which of the highlighted elements are siblings?	Clusters	Score	0.864	21.662	5	6.070 E-04 ∗∗∗
Training	3	11	Which two of the highlighted elements are more closely clustered?	Clusters	N/A				
Training	3	12	How many visually distinct clusters do you see in this visualization?	N/A	N/A				
Timed	3	13	Which two of the highlighted elements are more closely clustered?	Siblings	Score	0.738	29.499	5	1.850 E-05 ∗∗∗
Timed	3	14	Which two of the highlighted elements are more closely clustered?	Leaves	Score	0.275	12.775	5	2.558 E-02 ∗
Timed	3	15	Which two of the highlighted elements are more closely clustered?	Clusters	Score	0.352	9.539	5	8.941 E-02 ·
Timed	3	16	Which of the highlighted elements is least similar to its neighbors?	Clusters	Score	0.283	13.726	5	1.745 E-02 ∗
Timed	3	17	How many visually distinct clusters do you see in this visualization?	N/A	Value	8.794	31.138	5	8.796 E-06 ∗∗∗

Mechanical Turk User Study Analysis

Unboxing Cluster Heatmaps by Engle, Whalen, Joshi, and Pollard
6th Symposium on Biological Data Visualization (BioVis), October 23, 2016
Visualization and Graphics Lab, Department of Computer Science, University of San Francisco
Pollard Group, Gladstone Institutes, University of California San Francisco

Technique Pairs	Q13				Q14				Q15				Q16				Q17			
	est	err	t	p	est	err	t	p	est	err	t	p	est	err	t	p	est	err	t	p
Cluster Heatmap, Gapmap	1.056	0.542	1.949	0.361	-0.274	0.503	-0.546	0.994	-0.186	0.497	-0.374	0.999	0.916	0.531	1.726	0.511	3.121	0.945	3.303	0.014 *
Cluster Heatmap, Circle Packing	-0.065	0.504	-0.128	1.000	0.657	0.525	1.252	0.808	0.658	0.512	1.286	0.792	0.405	0.601	0.675	0.984	-3.745	0.960	-3.901	0.002 **
Cluster Heatmap, Sunburst	2.708	0.811	3.338	0.010 *	-1.191	0.571	-2.086	0.290	-1.159	0.537	-2.160	0.256	-0.292	0.579	-0.504	0.996	0.205	0.952	0.215	1.000
Cluster Heatmap, Radial Dendrogram	2.269	0.702	3.232	0.015 *	-2.032	0.698	-2.912	0.041 *	-1.159	0.537	-2.160	0.256	-0.965	0.662	-1.457	0.688	3.003	0.960	3.128	0.025 *
Cluster Heatmap, Force Directed Tree	1.504	0.573	2.623	0.087 ·	-0.903	0.533	-1.694	0.530	-1.041	0.524	-1.989	0.348	0.550	0.529	1.040	0.903	3.455	0.938	3.683	0.004 **
Gapmap, Circle Packing	0.992	0.546	1.817	0.443	0.383	0.538	0.711	0.980	0.473	0.516	0.916	0.942	1.322	0.581	2.275	0.202	-0.624	0.960	-0.650	0.987
Gapmap, Sunburst	1.652	0.838	1.972	0.347	-0.916	0.583	-1.571	0.612	-0.973	0.541	-1.801	0.464	-1.208	0.559	-2.163	0.252	-2.917	0.952	-3.063	0.030 *
Gapmap, Radial Dendrogram	1.213	0.732	1.656	0.550	-1.758	0.708	-2.483	0.126	-0.973	0.541	-1.801	0.464	-1.881	0.644	-2.921	0.040 *	-0.118	0.960	-0.123	1.000
Gapmap, Force Directed Tree	-0.448	0.610	-0.734	0.977	0.629	0.546	1.151	0.857	0.856	0.528	1.622	0.583	0.366	0.506	0.724	0.979	-0.333	0.938	-0.355	0.999
Circle Packing, Sunburst	2.644	0.814	3.248	0.014 *	-0.533	0.603	-0.885	0.949	-0.501	0.555	-0.903	0.946	0.113	0.625	0.181	1.000	-3.540	0.967	-3.660	0.004 **
Circle Packing, Radial Dendrogram	2.204	0.705	3.127	0.021 *	-1.375	0.724	-1.899	0.397	-0.501	0.555	-0.903	0.946	-0.560	0.703	-0.796	0.968	-0.742	0.975	-0.761	0.974
Circle Packing, Force Directed Tree	1.440	0.577	2.495	0.119	-0.246	0.567	-0.433	0.998	-0.383	0.542	-0.707	0.981	0.956	0.579	1.650	0.561	-0.290	0.953	-0.305	1.000
Sunburst, Radial Dendrogram	0.439	0.949	0.463	0.997	0.842	0.758	1.110	0.875	0.000	0.577	0.000	1.000	0.673	0.685	0.983	0.922	-2.798	0.967	-2.893	0.048 *
Sunburst, Force Directed Tree	1.204	0.859	1.402	0.716	-0.288	0.610	-0.472	0.997	-0.118	0.565	-0.208	1.000	-0.842	0.557	-1.513	0.653	-3.250	0.945	-3.438	0.009 **
Force Directed Tree, Radial Dendrogram	0.765	0.756	1.011	0.910	-1.129	0.730	-1.547	0.628	-0.118	0.565	-0.208	1.000	-1.515	0.642	-2.359	0.169	-0.452	0.953	-0.474	0.997

Mechanical Turk Post-Hoc Analysis

Unboxing Cluster Heatmaps by Engle, Whalen, Joshi, and Pollard
6th Symposium on Biological Data Visualization (BioVis), October 23, 2016
Visualization and Graphics Lab, Department of Computer Science, University of San Francisco
Pollard Group, Gladstone Institutes, University of California San Francisco