RnavGraph Interactive visual clustering

Adrian Waddell and Wayne Oldford

SSC June 14, 2011

WATERLOO

uwaterloo.ca

p values on each of n individuals

Challenge

p values on each of n individuals

data can have a complex structure



An end of the sector of the se







Challenge

p values on each of n individuals

data can have a complex structure



▶ *n*, or *p*, or both can be very large

Why Visualization?



- patterns, relations, textures,...
- recognize structure
- discover structure
- data analysis objectives
 - relations, clusters, classes, outliers, diagnostics, unusual structure, ...



Dealing with high dimensionality ightarrow large ho

- visually, we are constrained to small p
 - ▶ locations: *p* < 4
 - use color, shape, texture, movement,...
- large number of low dimensional views
 - $\binom{p}{d}$ *d*-dimensional views
 - How? Which ones? And, in what order?

Proposed Approach

- reduce numbers
 - p overall dimensionality
 - $\binom{p}{d}$ subset of all low-d views
- interactive/tailored low-d views
- connect low-d views via interactive navigation graphs

10.0	1-1	14.6	1.4	10.0	340	10×10	10.0	0.0	a . (20.0	24.6	86	2070	-	-	の変	19.8	-		30	13.4	30.0	13.0	150	10.0	100	a sa	43.0	1-1	Pres P	**	11-2	2-10	De .	100	De.	114	10	De .	1.4	114	114	11.6
8-16	3-8	10.0	1.1	100	24.6	0.0	B- B	000	8.0	20	0.0	24.6	820	10.0	15.0	84	10	15.0	No.	19.4	15 a	10.0	13.0	No.	15 m	10	asa a	020		10	10	14	15	12.0	1 × 1	10	32(1	900	11-	100	11.4	110	11.4
20.1	-	-		20.7	26		8-1	002	-	20	20.0	20.0	-	1.4	19	1	-		34	15 0	10	33	150	10	Be	10.5	150	205	-	-	-	1	-	-	-	-		-	1-1		1.4		10
8.0	-	-	-	20	36	20	-	00	-	20	30	20	8.0	1.0	-	100	-	-	100	105	5.0		10	105	80	10.5	10	20		-	-	1	10	-	-	-		-	-	-	12	100	
8.0	-	-	-	20	20	-	-	007	-	ac	30	204	8.0	10	-	10.5	-	-	-	100	-		100	8.6	88	20	Se la	-	-	-	-	1	-	10	10	-	-	12	1 1	-	-	-	-
0.0	10	20		20	26	20	() R.B	20	-	20	20	100	9.0	10	-	8.6		-	100	1.0		200	50	8.0	9.0	10	10	100	-	-	1	12 10	3		1	10	-	-	1 10	-	44	10	8.00
6 80	100	10	1.4	10	10	10		100	30	20	6 20	0.0	10	-	10	10	10	-	-	1 10	-	8	100	10		20	10	0 0	-	1	-	1 1	10	1	1	10	(1	10	1 1	1	40	-	8
6 10	10	10	1 10	() 10	0.00	1 20	10	0 8	30	00	6 20	20	20	44	10	0 0	50	10	1 10	1 20	-	100	10	6 9	1	20	1	9 9	10	1	1	3 11-	10	-	1	100	10	10	10	1			100
6 20	10	10	100	100	100	10 10	100		20	20	10.00	20	100		100	100	10	100	1	100	-	100	10	-	100	100	100	20	1	-	-	1	in la	-	101	100	70. 4	-			2	10	-
000	10	il.	34	10.0	1	-	-10	100	20	10	10/	10	-0		10	3	-	3		50	御	3	1		-	(<u>3</u>)	1	-	-	100	104	No.	in the	in .	100	200	10.4	104	-	200	1. All	10.	20
1) 2	-10	36		1	19.4	9.6	-0	-	1	(e)	-0	-0	94	-	10	8	1	3	3	-	1	8	10		- 22		3	19.4	10	0.e	10.4	10.	9.6	20	20	2.0	104	100		104	No.	114	90
10.	9	-18	10.0	0	- (1)	.0		10	1	-12	in	.0		-	3	0	16		3	19.0	J.	3	3		-	6)	-	9.0	10	Det	104	11-2	1	1	104	1.0	94	100		1)4	10	104	100
0 . ((18 × (1)	10.8	8 × (1)	000	2.0	300	E all	10.0	20	20.0	10 + G	20.0	10.0	の	-	100	T'an	-	の	200	19.00	10	8 - C	00.3	-	10	10 a	10	114	1) c	10-1	Pril 1	100	1	100	100	1) < B	1) c 8		1 × (1	100	10	10.5
9+6	10.00	10.00		20.0	1.0	8.1	10.0	B-B	20	20.0	84	10×10	20	-	10	15.0	10	10	18.0	19.0	「日日	100	19.00	0 C		10	の	1 cal	1-10	100	10	P.u.	11	De B	De a	2.0	1)0	1) 50	-	100	De B	10.0	Deg
9.6	10.0	8.6	9-8	9.0	8×6	a . (10 m	100	10 m	2.0	810	0.0	10 th	の	13.8	10 a	0.0	15.0	100	19.00	12.0	の行	150	0.1.3	20	10.00	18.6	100	100		10	123	14	11	10.4	2.4	10 4	8-10	10	10.4	100	1 1 1 H	Deg
8-10	10	20.70	10	100	10.0	20.0	8 2 (8.0	202	20.0		201	14.0	13.0	13.0	19.6	10	18.0	100	Be	Be	18.8	Be	2013	Be	3.8	B.	100	10	T	14	13	1	12	10	210	10	De B	10	12	10	100	0cg
-	-	-	-	20.10	-	20	-	-	a.c.	20	24	20	8-1	-	84	19.60	-	10	8.	19.6	15.0	1 a	Be	80	B.ª	19.51	18 a			75	-	13	1	1	-	1.4	-	-	-		1.4	-	10
-		-	-	20.7	-	20		-	80	20	30.1	20	8.0	100	10	19.6	-	-	100	80	100	-	19.0	9.0	19	-	100		-	3	4	5	5	-	1.0		-	-		-10	-	-	1.2
-	-	-	-	20	-	20		8.0	20	20	R.C.	20	20			20		-	-	20	10	10	80	20	-	-	-	-	-	-	-	E	-	10	10	10	-	-	-	-	-	10	-
	10	-	10	10	8.8	20	-		20	20	30	20	100	-		-		1	4.6	1 0	10		20	0.0	10	-	100	-		-	10	1	1		4 10	11 12	1	1			10	1	1
-	1 2 10	12	10	(E) 40	-	100	1	1 10	8	120	a.c.	10	100	-	0 0	10	-	10	-	(0 0	1 10	10	80	30	20	20	-	8	10	-	100	1 1	-	1	10	10 2	10		-	10	-	-	100
-	10	10	10	10	10 10	8	6 9	(1)	30	8	20	100	(1)	-	600	10	-	10	10	10 m	5	30	10	8	20	0.0	10		-	-	10.4	104	100	10	1 1) e	10	10	10	-	10	-	200	10.00
1	10	1).	10.	100	2 (R) #	-	20 94	200	4	100	20	-	2	-	3	1	-	-	3	Si al	(4)	(A)	-	3	-	-	-	10	100	De	104	100	in a	1	104	200	200	30	1		*	10	10
-	14	-	- 6	1).	9.4	.0	94	9.0	1	6	E.	.0	34	8	1	3		3	3	13	1	10	19.6		10	6	100	10	10	De-	De.	Be.	10	10	1	00	De.	36	-	-	No.	10	94





> $28 \cdot 20 = 560$ dimensions



- ▶ $28 \cdot 20 = 560$ dimensions
- explore via low dimensional spaces



- > $28 \cdot 20 = 560$ dimensions
- explore via low dimensional spaces
- Using LLE: local linear embedding
- k = 12 neighbors
- reduce to 5 dimensions

 $(x_1, x_2, x_3, x_4, x_5)$







Deactivate complementary points





Zoom out, reactivate, resize, pan







► 3*d* rigid rotation





SSC, RnavGraph, June 14, 2011









$$\binom{5}{2} = 10$$



Each node is a 2d view

$$\binom{5}{2} = 10$$

► Each edge is a 3*d* transition



Each node is a 2d view

$$\binom{5}{2} = 10$$

Each edge is a 3*d* transition



$$\binom{5}{2} = 10$$

- Each edge is a 3*d* transition
- You are here bullet



$$\binom{5}{2} = 10$$

- Each edge is a 3*d* transition
- You are here bullet
- This is a navigation graph (NavGraph)



$$\binom{5}{2} = 10$$

- Each edge is a 3*d* transition
- You are here bullet
- This is a navigation graph (NavGraph)

Move the bullet

2*d* view changes in response

Can stop anywhere and interact with the low-d view

3d and 4d transition graphs



complement of 3d transition graph yields 4d transition graph

3d and 4d transition graphs



complement of 3d transition graph yields 4d transition graph

3d and 4d transition graphs



complement of 3d transition graph yields 4d transition graph











4d transition is NOT a rigit rotation

Try the package yourself

- This is only a part of our R package
- Try it yourself
- Package is on CRAN: install.packages('RnavGraph')
- Install dependencies and suggested packages
- Read the vignette
- Try the demos

Papers

- Oldford and Waddell
 - Visual clustering of high-dimensional data by navigating low-dimensional spaces (ISI Dublin, 2011)
 - RnavGraph: A visualization tool for navigating through high dimensionaldata (ISI Dublin, 2011)
 - RnavGraph R package, available on CRAN
- Hurley and Oldford
 - Graphs as a navigational infrastructure for high dimensional data spaces (Comp Stats 2011)
 - Pairwise display of high dimensional information via Eulerian tours and Hamiltonian decompositions (JCGS, 2010)
 - Eulerian tour algorithms for data visualization and the PairViz R package (Comp Stats 2011)
 - PairViz R package, available on CRAN