

Carnegie Mellon CBD Alumni Lightning talk series

Arshi Arora

Research Biostatistician II

Department of Epidemiology and Biostatistics

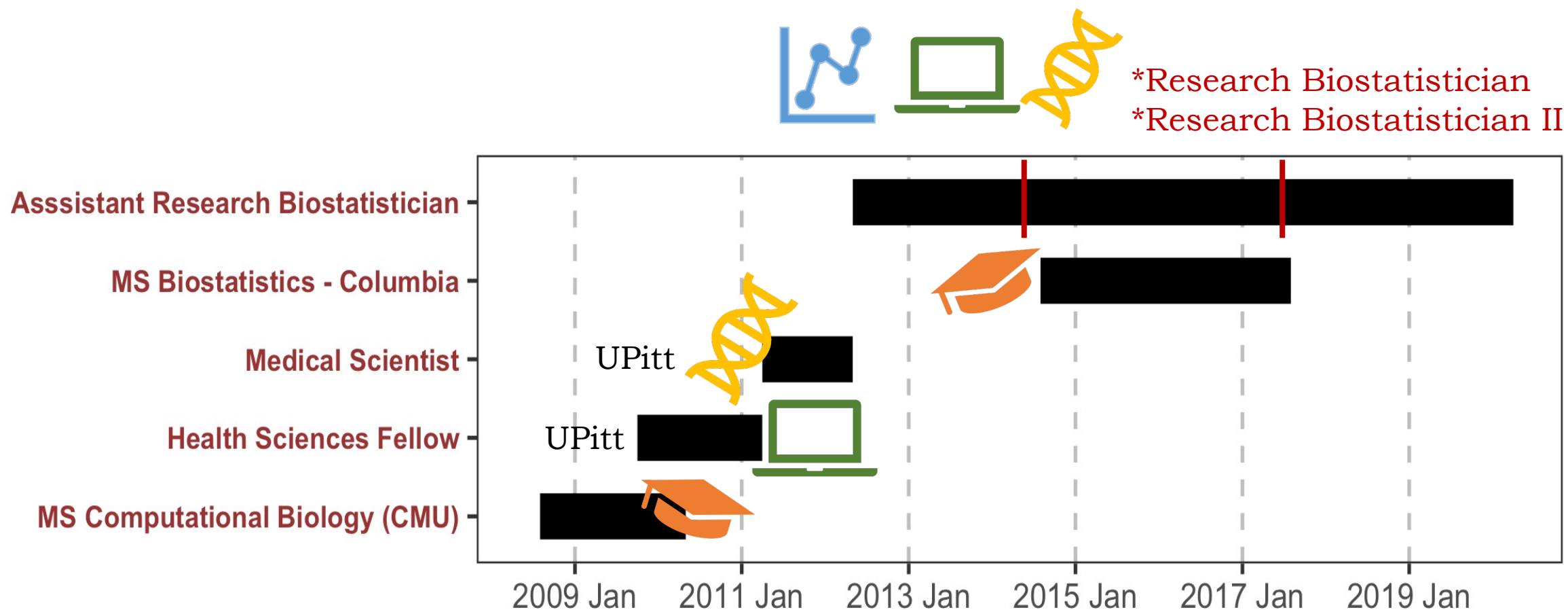
Memorial Sloan Kettering Cancer Center

04/07/2020



Memorial Sloan Kettering
Cancer Center™

Journey so far



Journey so far in few words

- Cancer genomics
 - Multi-omic integrative analysis (TCGA)
 - Tumor heterogeneity
 - Building prognostic models
 - Identifying risk groups
- Rmarkdown pipeline – Reproducibility
- Make your work available on GitHub, R package, perl scripts

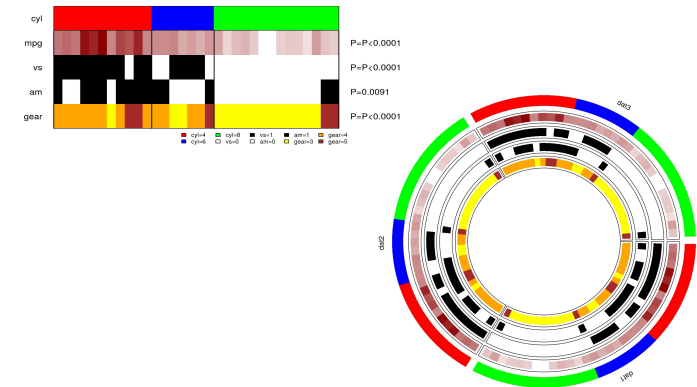
Current work

- *survClust*

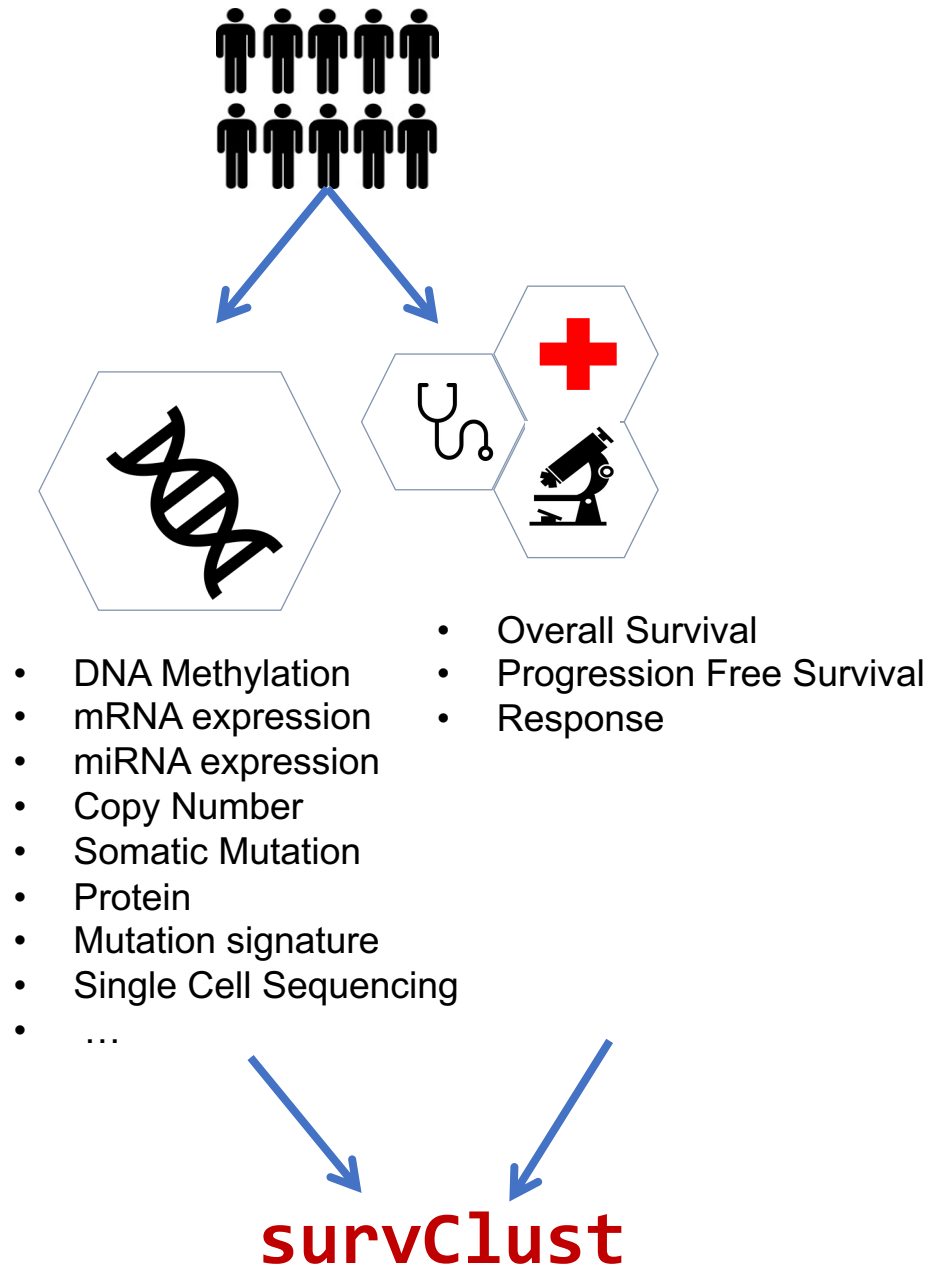
identification of clinically relevant genomic subtypes using outcome-weighted integrative clustering

- *Panelmap*

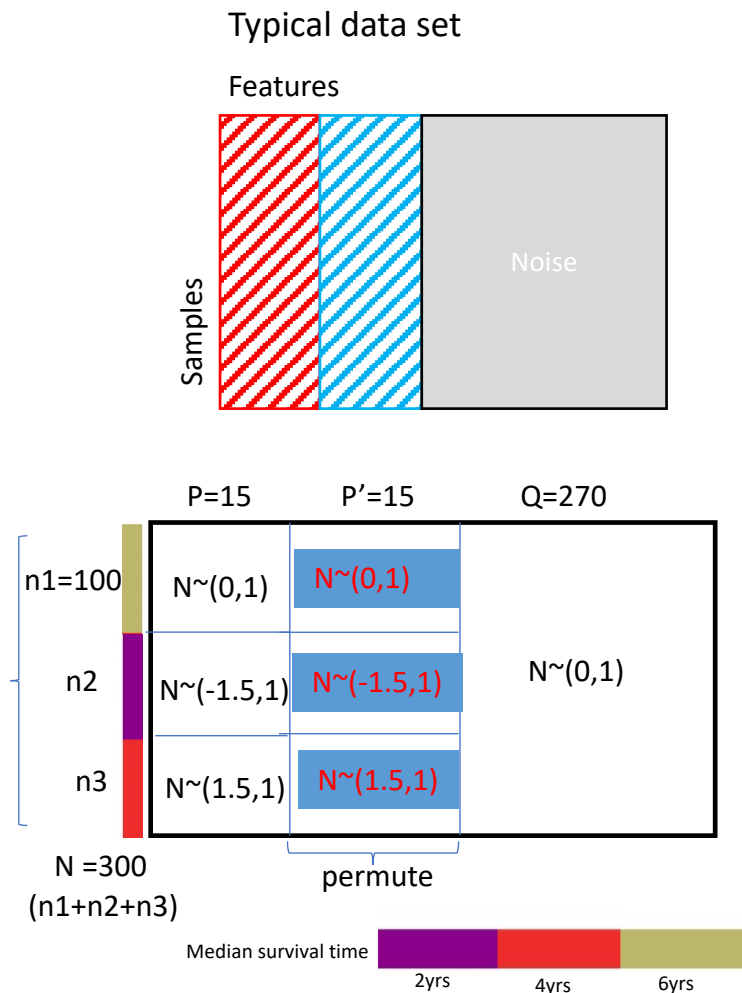
A data visualization tool for viewing clustered data and its differentiated features in single and multi-panel layout.



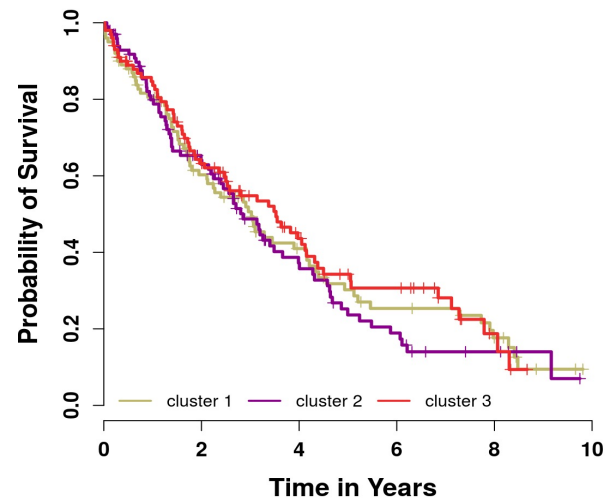
survClust



unsupervised vs supervised clustering via simulation

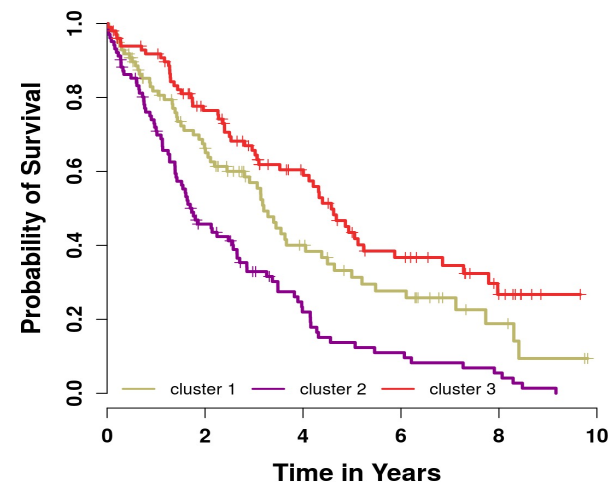


Unsupervised clustering solution*
K-means solution (k=3)



	Cluster 1	Cluster 2	cluster 3
Truth Cluster 1	29	34	38
Truth Cluster 2	71	0	0
Truth Cluster 3	0	66	62

survClust solution
survClust solution (k=3)



	Cluster 1	Cluster 2	cluster 3
Truth Cluster 1	100	0	0
Truth Cluster 2	0	100	0
Truth Cluster 3	0	0	100

* Unsupervised clustering solution was arrived by running *k-means* algorithm

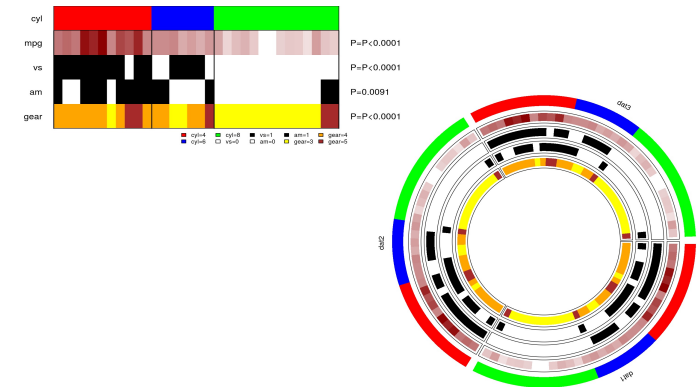
Current work

- *survClust*

identification of clinically relevant genomic subtypes using outcome-weighted integrative clustering

- *Panelmap*

A data visualization tool for viewing clustered data and its differentiated features in single and multi-panel layout.



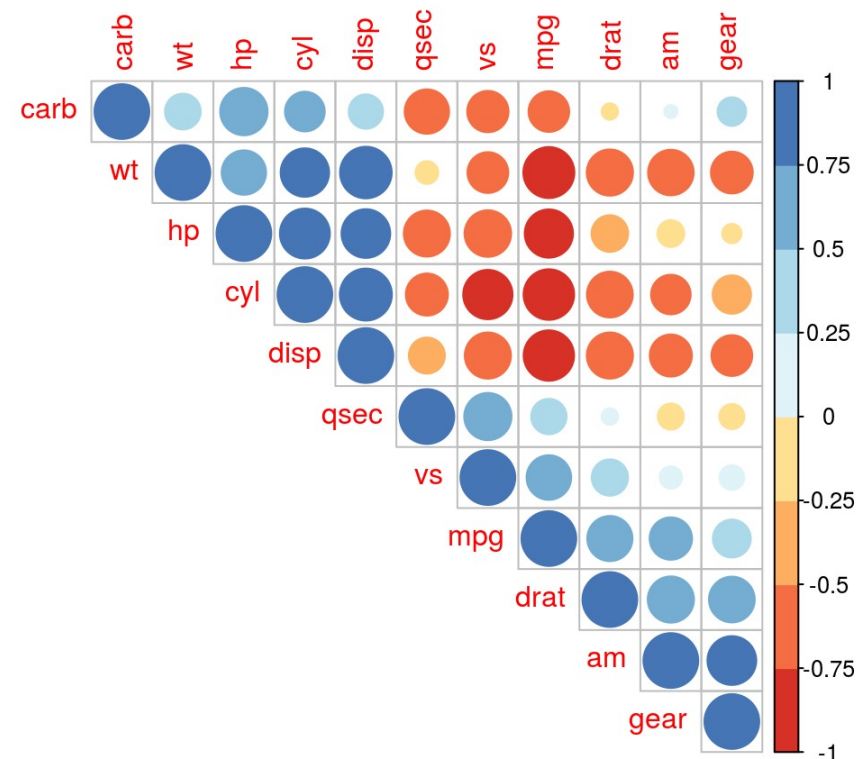
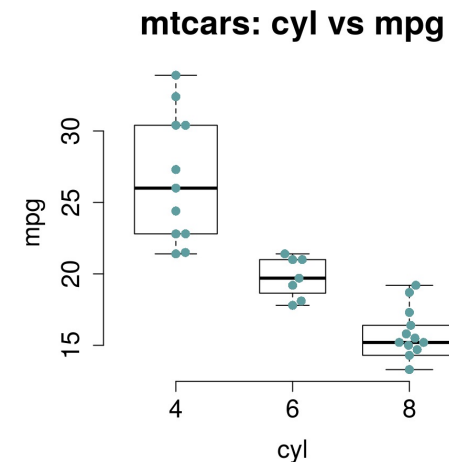
Panelmap - Motivation example

mtcars dataset

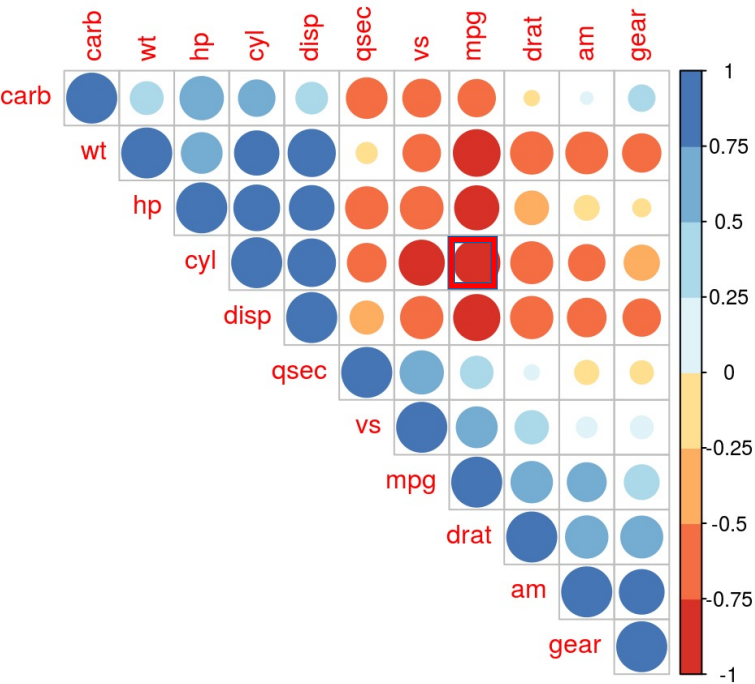
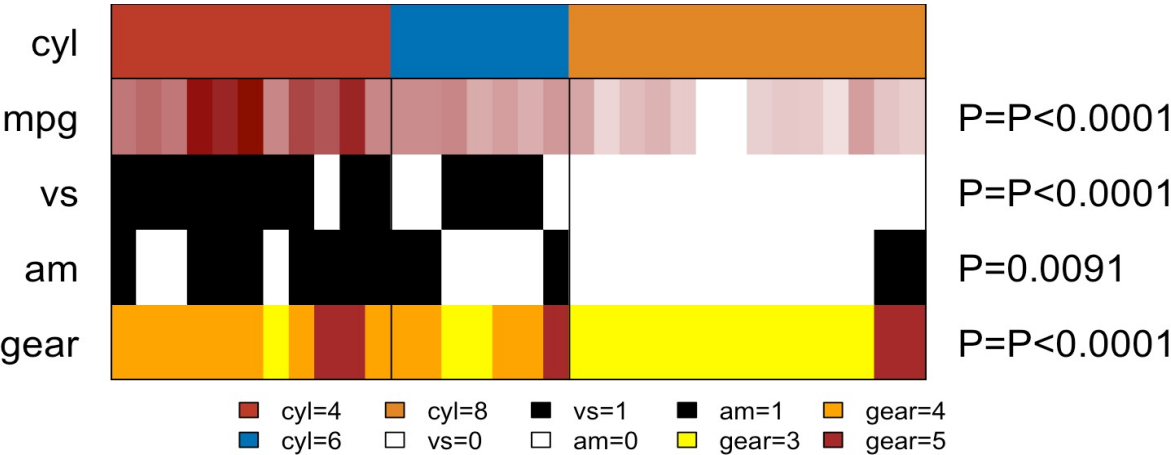
	mpg	cyl	disp	hp	drat	wt	qsec	vs	am	gear	carb
Mazda RX4	21.0	6	160	110	3.90	2.620	16.46	0	1	4	4
Mazda RX4 Wag	21.0	6	160	110	3.90	2.875	17.02	0	1	4	4
Datsun 710	22.8	4	108	93	3.85	2.320	18.61	1	1	4	1
Hornet 4 Drive	21.4	6	258	110	3.08	3.215	19.44	1	0	3	1
Hornet Sportabout	18.7	8	360	175	3.15	3.440	17.02	0	0	3	2
Valiant	18.1	6	225	105	2.76	3.460	20.22	1	0	3	1

vs = 0, V shaped engine
vs = 1, S shaped engine

am=0, automatic
am=1, manual

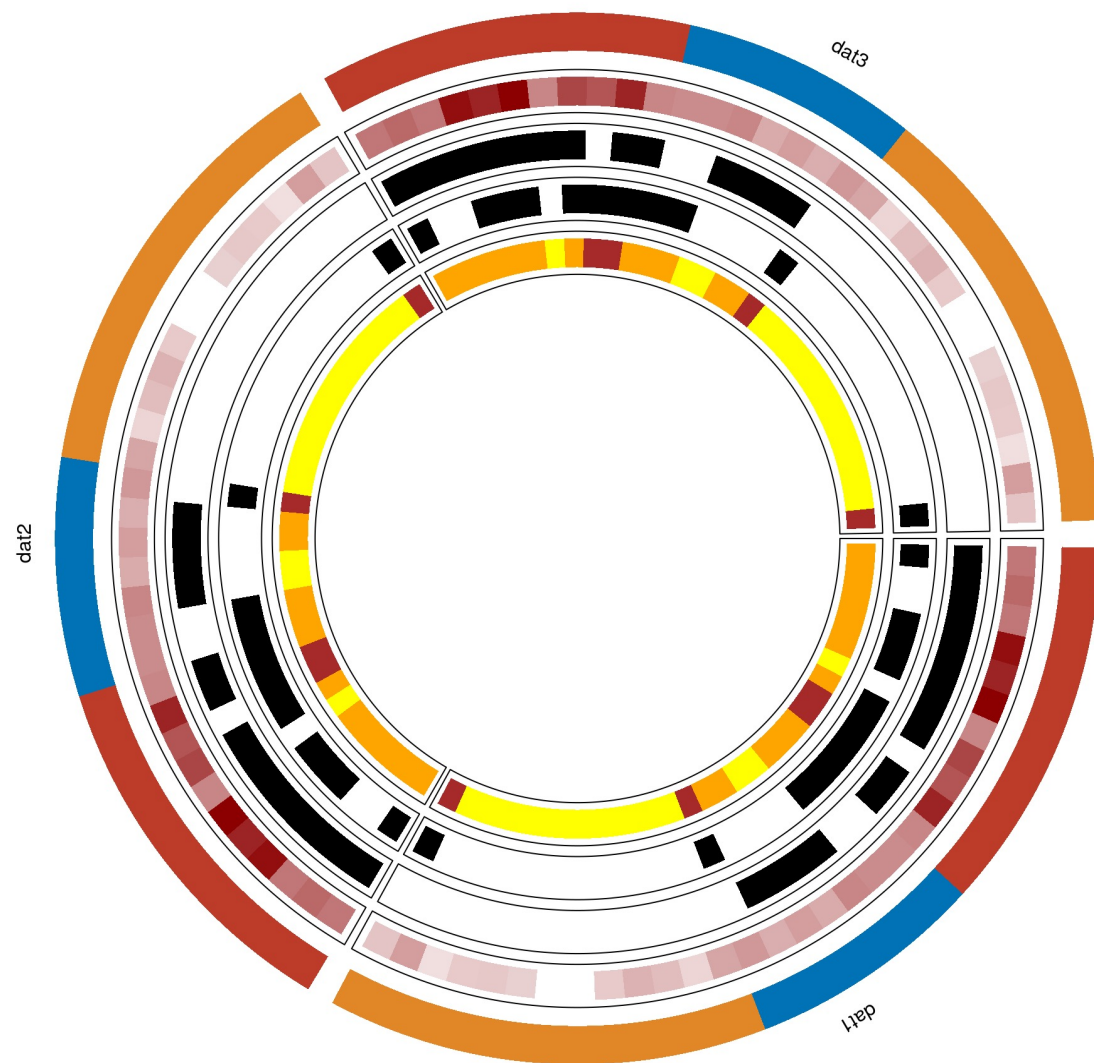
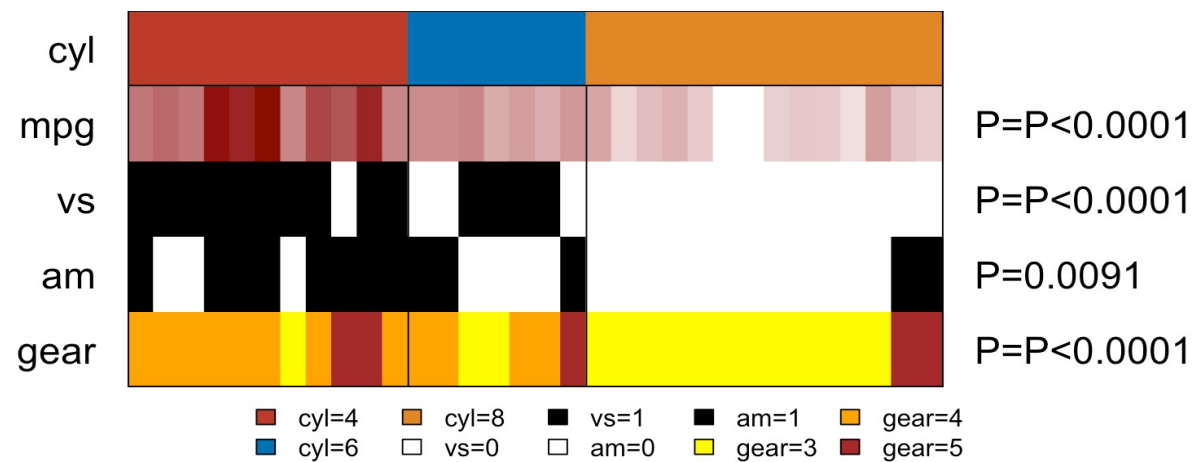


Introducing – panelmap on mtcars dataset



	4(n=11;34.38%)	6(n=7;21.88%)	8(n=14;43.75%)	RowTotal	pval
mpg	26[21.4-33.9]	19.7[17.8-21.4]	15.2[10.4-19.2]		$P<0.0001$
NA	0	0	0		
vs					$P<0.0001$
0	1(9%)	3(43%)	14(100%)	18(56.25%)	
1	10(91%)	4(57%)	0(0%)	14(43.75%)	
am					0.0091
0	3(27%)	4(57%)	12(86%)	19(59.38%)	
1	8(73%)	3(43%)	2(14%)	13(40.62%)	
gear					$P<0.0001$
3	1(9%)	2(29%)	12(86%)	15(46.88%)	
4	8(73%)	4(57%)	0(0%)	12(37.5%)	
5	2(18%)	1(14%)	2(14%)	5(15.62%)	

circomap - Multiple panelmaps



GitHub

- Panelmap - <http://github.com/arorarshi/panelmap>
- survClust - <http://github.com/arorarshi/survClust>

Reach out to me at

- arshiaurora@gmail.com
- GitHub – <http://github.com/arorarshi>