Fong Chun Chan | @fongchunchan | tinyheero.github.io

GENERATING HEATMAPS IN R

Objectives

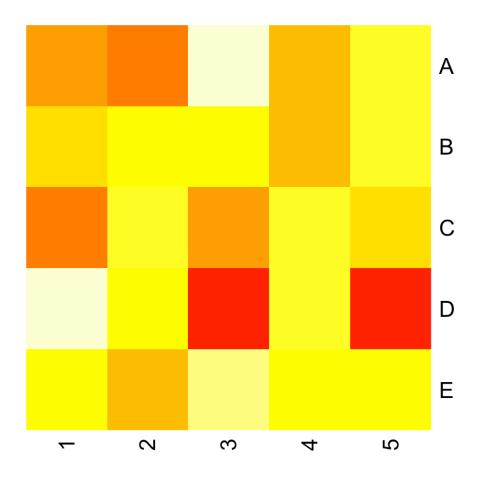
- Learn what heatmaps are
- Learn how to generate a basic heatmap in R
- Learn the importance of normalization and clustering in heatmap generation

Consider this 5x5 numeric matrix:

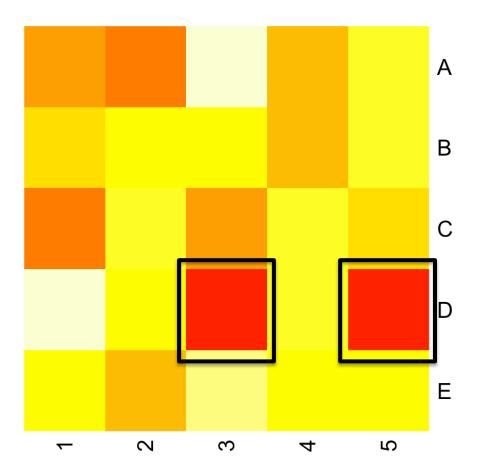
	1	2	3	4	5
Α	-0.6264538	-0.8204684	1.5117812	-0.0449336	0.9189774
В	0.1836433	0.4874291	0.3898432	-0.0161903	0.7821363
С	-0.8356286	0.7383247	-0.6212406	0.9438362	0.0745650
D	1.5952808	0.5757814	-2.2146999	0.8212212	-1.9893517
Е	0.3295078	-0.3053884	1.1249309	0.5939013	0.6198257

Can you identify the outliers/extreme values?

Consider this 5x5 numeric matrix:



Consider this 5x5 numeric matrix:



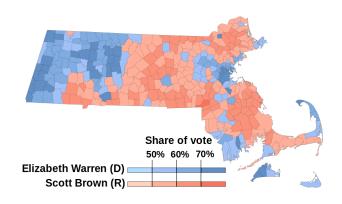
Consider this 5x5 numeric matrix:

	1	2	3	4	5
Α	-0.6264538	-0.8204684	1.5117812	-0.0449336	0.9189774
В	0.1836433	0.4874291	0.3898432	-0.0161903	0.7821363
С	-0.8356286	0.7383247	-0.6212406	0.9438362	0.0745650
D	1.5952808	0.5757814	-2.2146999	0.8212212	-1.9893517
E	0.3295078	-0.3053884	1.1249309	0.5939013	0.6198257

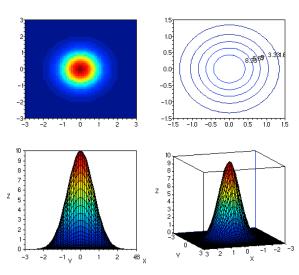
• Can you identify the outliers/extreme values?

- A heatmap is graphical representation of values by colors.
- They are often used for matrix data, but not restricted to this data format.

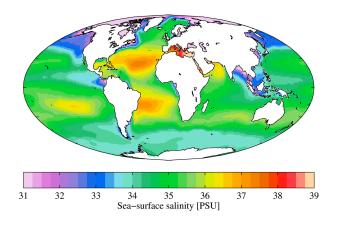
Heatmap Examples



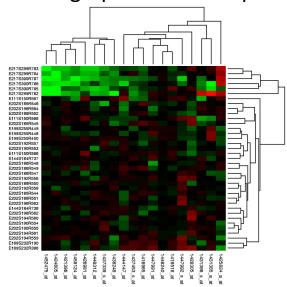
Choropleth Heatmap



Density Heatmap



Geographical Heatmap

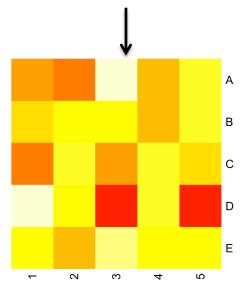


"Expression" Heatmap

Generating Heatmaps in R

 For this workshop, we will be generating heatmaps from matrix data:

	1	2	3	4	5
Α	-0.6264538	-0.8204684	1.5117812	-0.0449336	0.9189774
В	0.1836433	0.4874291	0.3898432	-0.0161903	0.7821363
С	-0.8356286	0.7383247	-0.6212406	0.9438362	0.0745650
D	1.5952808	0.5757814	-2.2146999	0.8212212	-1.9893517
E	0.3295078	-0.3053884	1.1249309	0.5939013	0.6198257



Coding Exercise #1

 Open the "heatmap-cluster.Rmd file, and go through the code chunks in section "Generating Heatmaps from Matrix Data in R"

Coding Exercise

- By the end of this coding exercise, you should be able to:
 - 1. Use the image() function to generate the most basic heatmap
 - Recognize the "quirks" of image()
 - 3. How heatmap() and d3heatmap() are better functions for heatmap generation

Problems with "Out of Box" Heatmap

 So far we have taken a matrix and directly converted it into a heatmap using the image() function.

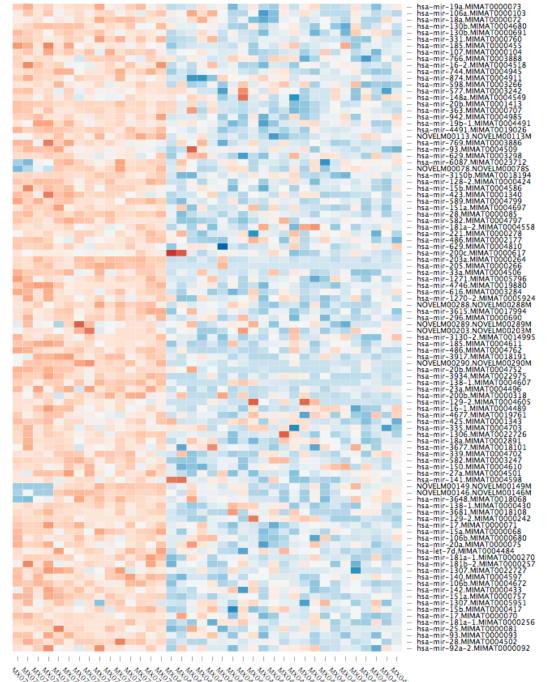
Does this "out of box" solution always work?

miRNA Expression Matrix

(Lim et al. 2015; Comprehensive miRNA sequence analysis reveals survival differences in diffuse large B-cell lymphoma patients)

Low

High



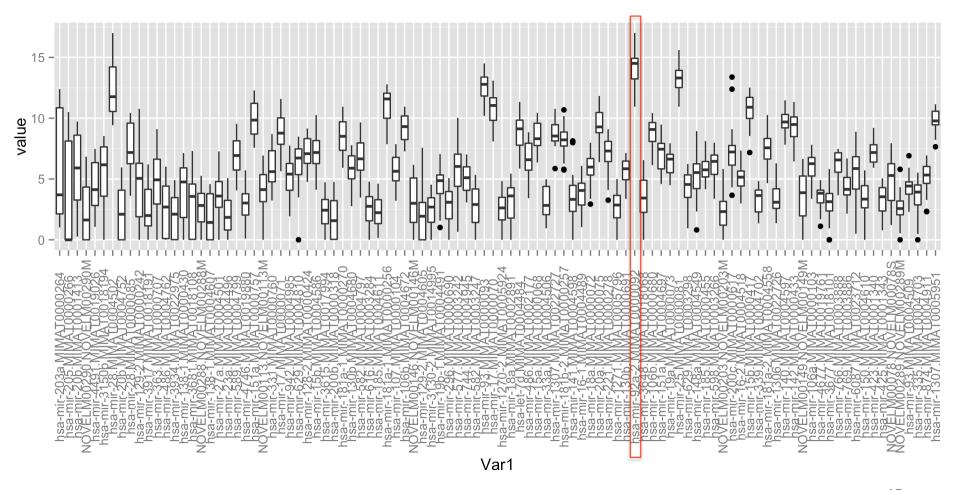
Samples

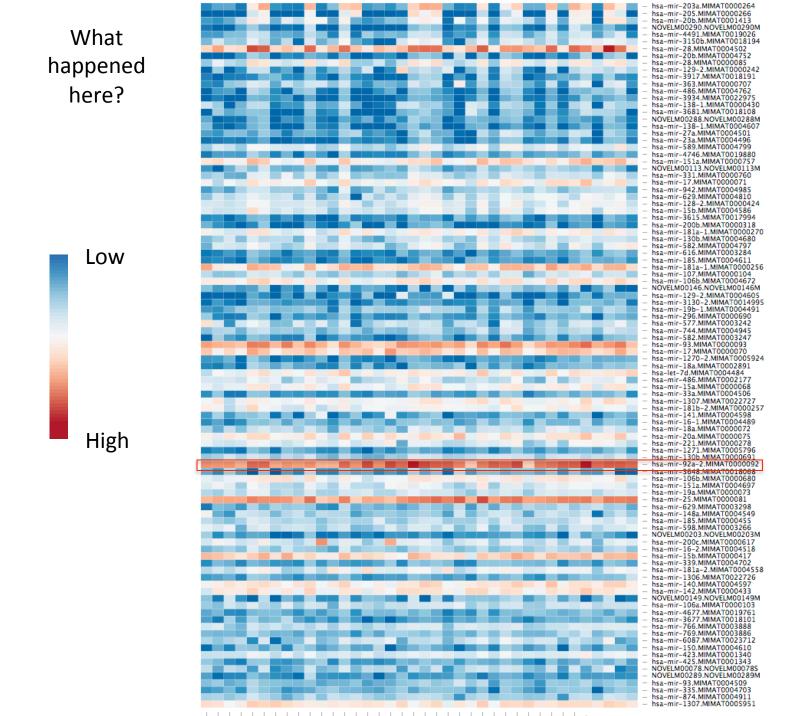
Coding Exercise #2

 In this exercise, we will take an "expression" matrix and attempt to produce a heatmap from it.

 Go through the code chunks in section "Heatmap Normalization"

Feature Value Distribution



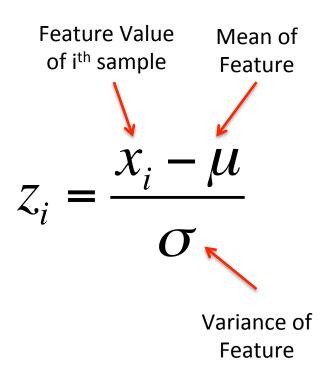


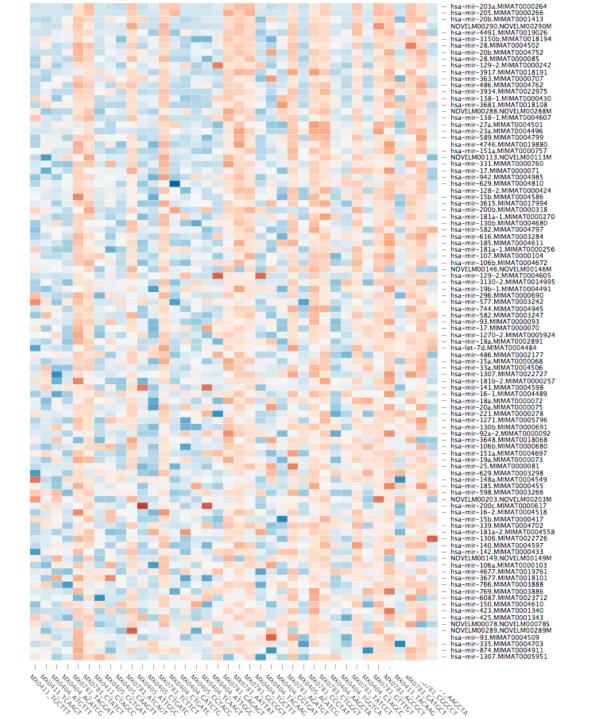
Importance of Normalization

- Each feature has its own distribution; In other words, the features are NOT on the same scale.
- The differences between the features outweighs differences between samples
 - Encoding of values into colors gives "striped" results.
- Need to normalize the features such that they are on the same scale

Z-score Normalization

 One of the most common normalization technique is the Z-score normalization





Bonus Question(s) / Food for Thought

 How could you perform a different type of normalization?

Should you normalize the columns too?

hsa-mir-203a.MIMAT0000264 hsa-mir-205.MIMAT0000266 hsa-mir-20b.MIMAT0001413 NOVELM00290.NOVELM00290M hsa-mir-4491.MIMAT0019026 hsa-mir-3150b.MIMAT0018194 hsa-mir-28.MIMAT0004502 hsa-mir-20.MIMAT0004752 hsa-mir-28.MIMAT000085 hsa-mir-129-2.MIMAT0000242 hsa-mir-3917.MIMAT0018191 hsa-mir-363.MIMAT000707 hsa-mir-486.MIMAT0004762 hsa-mir-3934.MIMAT0022975 hsa-mir-138-1.MIMAT0000430 hsa-mir-3681.MIMAT0018108 NOVELM00288.NOVELM00288M hsa-mir-138-1.MIMAT0004607 hsa-mir-27a.MIMAT0004501 hsa-mir-23a.MIMAT0004496 hsa-mir-589.MIMAT0004799 hsa-mir-4746.MIMAT0019880 hsa-mir-151a.MIMAT0000757 NOVELM00113.NOVELM00113M hsa-mir-331.MIMAT0000760 hsa-mir-17.MIMAT0000071 hsa-mir-942.MIMAT0004985 hsa-mir-629.MIMAT0004810 hsa-mir-128-2.MIMAT0000424 hsa-mir-15b.MIMAT0004586 hsa-mir-3615.MIMAT0017994 hsa-mir-200b.MIMAT0000318 hsa-mir-181a-1.MIMAT0000270 hsa-mir-130b.MIMAT0004680 hsa-mir-130b.MIMAT0004797 hsa-mir-582.MIMAT0004797 hsa-mir-616.MIMAT0003284 hsa-mir-185.MIMAT0004611 hsa-mir-181a-1.MIMAT0000256 hsa-mir-107.MIMAT0000104 hsa-mir-107.MIMAT0000104 hsa-mir-106b.MIMAT0004672 NOVELM00146.NOVELM00146M0 hsa-mir-129-2.MIMAT0004605 hsa-mir-1310-2.MIMAT001490 hsa-mir-196.MIMAT000491 hsa-mir-296.MIMAT0000530 hsa-mir-744.MIMAT0003242 hsa-mir-744.MIMAT0003247 nsa-mir-382.MIMA10000293 hsa-mir-93.MIMAT0000093 hsa-mir-17.MIMAT0000070 hsa-mir-1270-2.MIMAT0005924 hsa-mir-18a.MIMAT0002891 hsa-let-7d.MIMAT0004484 hsa-mir-486.MIMAT0002177 hsa-mir-15a.MIMAT000068 hsa-mir-33a.MIMAT0004506 hsa-mir-1307.MIMAT0022727 hsa-mir-181b-2.MIMAT0000257 hsa-mir-141.MIMAT0004598 hsa-mir-16-1.MIMAT0004489 hsa-mir-18a.MIMAT0000072 hsa-mir-20a.MIMAT0000075 hsa-mir-221.MIMAT0000278 hsa-mir-1271.MIMAT0005796 hsa-mir-130b.MIMAT0000691 hsa-mir-92a-2.MIMAT0000092 hsa-mir-3648.MIMAT0018068 hsa-mir-106b.MIMAT0000680 hsa-mir-151a.MIMAT0004697 hsa-mir-19a.MIMAT0000073 hsa-mir-25.MIMAT0000081 hsa-mir-629.MIMAT0003298 hsa-mir-148a.MIMAT0004549 hsa-mir-185.MIMAT000455 hsa-mir-598.MIMAT0003266 NOVELM00203.NOVELM00203M hsa-mir-200c.MIMAT0000617 hsa-mir-16-2.MIMAT0004518 hsa-mir-15b.MIMAT0000417 hsa-mir-339.MIMAT0004702 hsa-mir-181a-2.MIMAT0004558 hsa-mir-1306.MIMAT0022726 hsa-mir-140.MIMAT0004597 nsa-mir-142.MIMAT00004597 hsa-mir-142.MIMAT0000433 NOVELM00149.NOVELM00149M hsa-mir-106a.MIMAT0000103 hsa-mir-4677.MIMAT0019761 hsa-mir-3677.MIMAT0018101 hsa-mir-766.MIMAT0003888 hsa-mir-769.MIMAT0003886 hsa-mir-6087.MIMAT0023712 hsa-mir-150.MIMAT0004610 hsa-mir-423.MIMAT0001340 hsa-mir-425.MIMAT0001343 NOVELM00078.NOVELM000785 NOVELM00289.NOVELM00289M hsa-mir-93.MIMAT0004509 hsa-mir-335.MIMAT0004703 hsa-mir-874.MIMAT0004911 hsa-mir-1307.MIMAT0005951

Similar Samples?

Heatmap Clustering

• Clustering is a machine learning technique to "group" together (i.e. cluster) similar things.

Data can be clustered for both rows and columns

Coding Exercise #3

 Go through the code chunks in section "Heatmap Clustering"

- Bonus Question:
 - How could you try different clustering methods?

Conclusions

- By the end of this workshop, hopefully you now:
 - Under what heatmaps
 - How to generate basic heatmaps in R
 - How to normalize and clustering your data to produce a nice and final output
- Future Areas of Interest
 - Exploring different forms of clustering
 - Extracting the dendrograms
 - Adding annotation tracks for rows and samples