

Fong Chun Chan | @fongchunchan | [tinyheero.github.io](https://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

# What is a Heatmap?

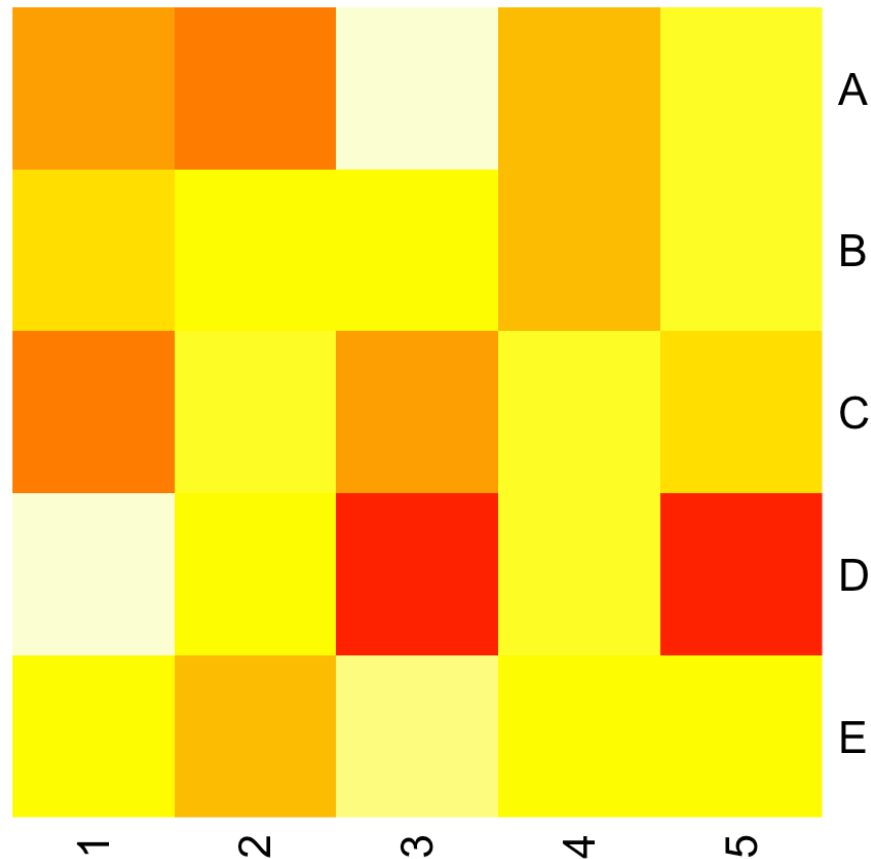
- Consider this 5x5 numeric matrix:

	1	2	3	4	5
A	-0.6264538	-0.8204684	1.5117812	-0.0449336	0.9189774
B	0.1836433	0.4874291	0.3898432	-0.0161903	0.7821363
C	-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?

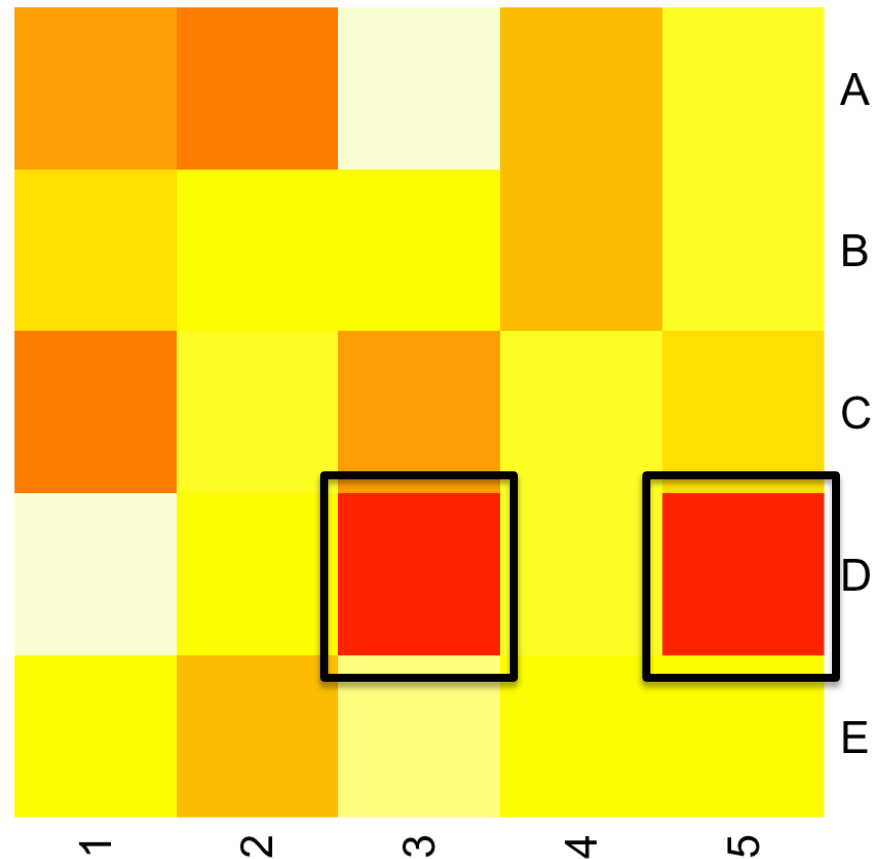
# What is a Heatmap?

- Consider this 5x5 numeric matrix:



# What is a Heatmap?

- Consider this 5x5 numeric matrix:



# What is a Heatmap?

- Consider this 5x5 numeric matrix:

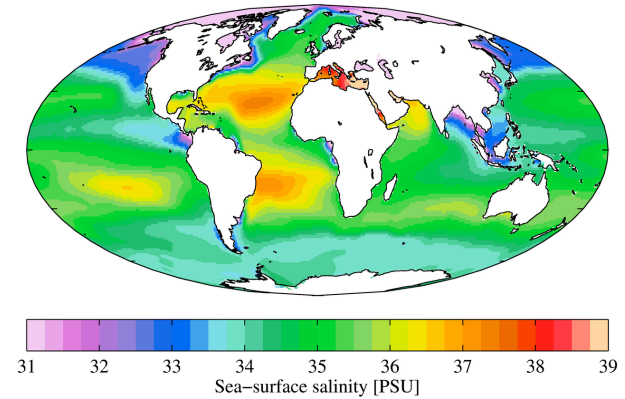
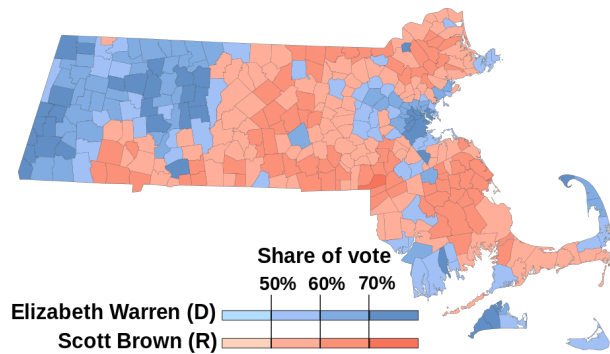
	1	2	3	4	5
A	-0.6264538	-0.8204684	1.5117812	-0.0449336	0.9189774
B	0.1836433	0.4874291	0.3898432	-0.0161903	0.7821363
C	-0.8356286	0.7383247	-0.6212406	0.9438362	0.0745650
D	1.5952808	0.5757814	-2.2146999	0.8212212	-1.9893517
E	0.3295078	-0.3053884	1.1249309	0.5939013	0.6198257

- Can you identify the outliers/extreme values?

# What is a Heatmap?

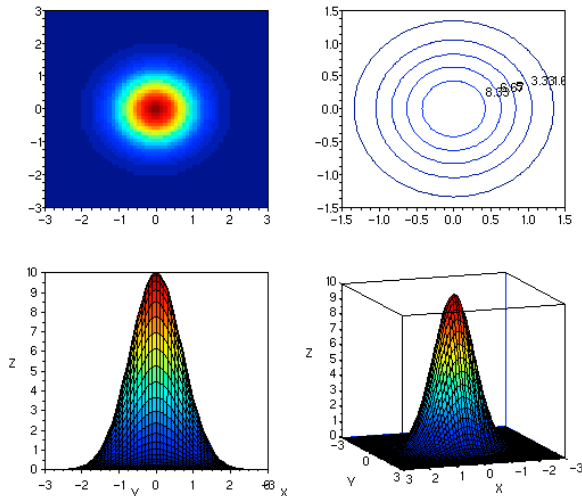
- A heatmap is graphical representation of values by colors.
- They are often used for matrix data, but not restricted to this data format.

# Heatmap Examples

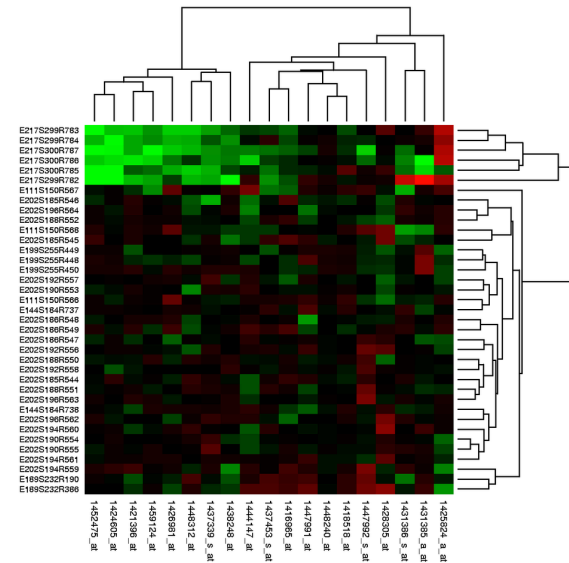


Choropleth Heatmap

Geographical Heatmap



Density Heatmap



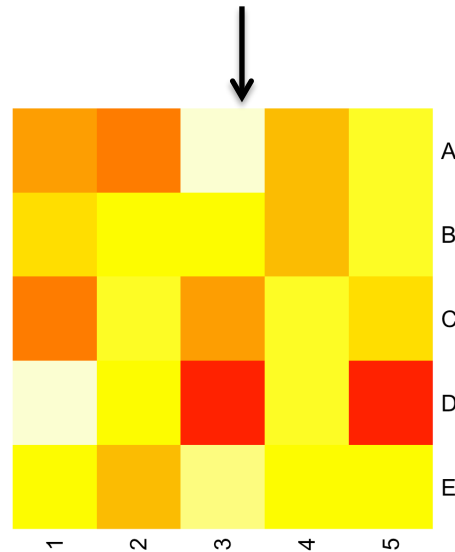
“Expression” Heatmap



# Generating Heatmaps in R

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

	1	2	3	4	5
A	-0.6264538	-0.8204684	1.5117812	-0.0449336	0.9189774
B	0.1836433	0.4874291	0.3898432	-0.0161903	0.7821363
C	-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

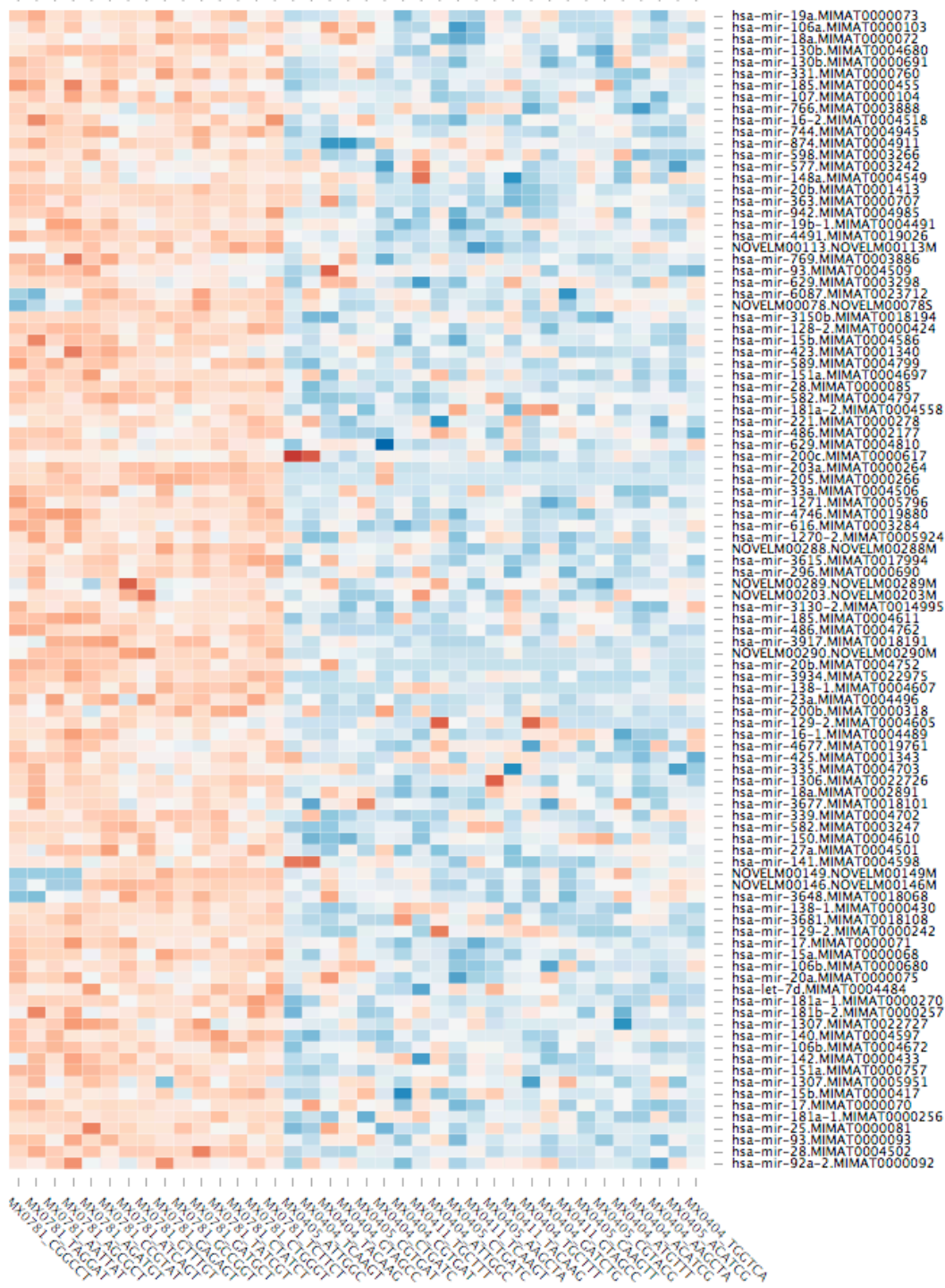
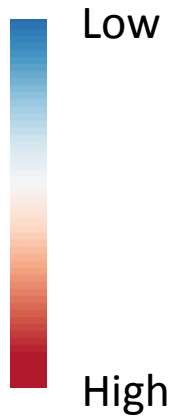
- By the end of this coding exercise, you should be able to:
  1. Use the `image()` function to generate the most basic heatmap
  2. Recognize the “quirks” of `image()`
  3. How `heatmap()` and `d3heatmap()` are better functions for heatmap generation

# 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?

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

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



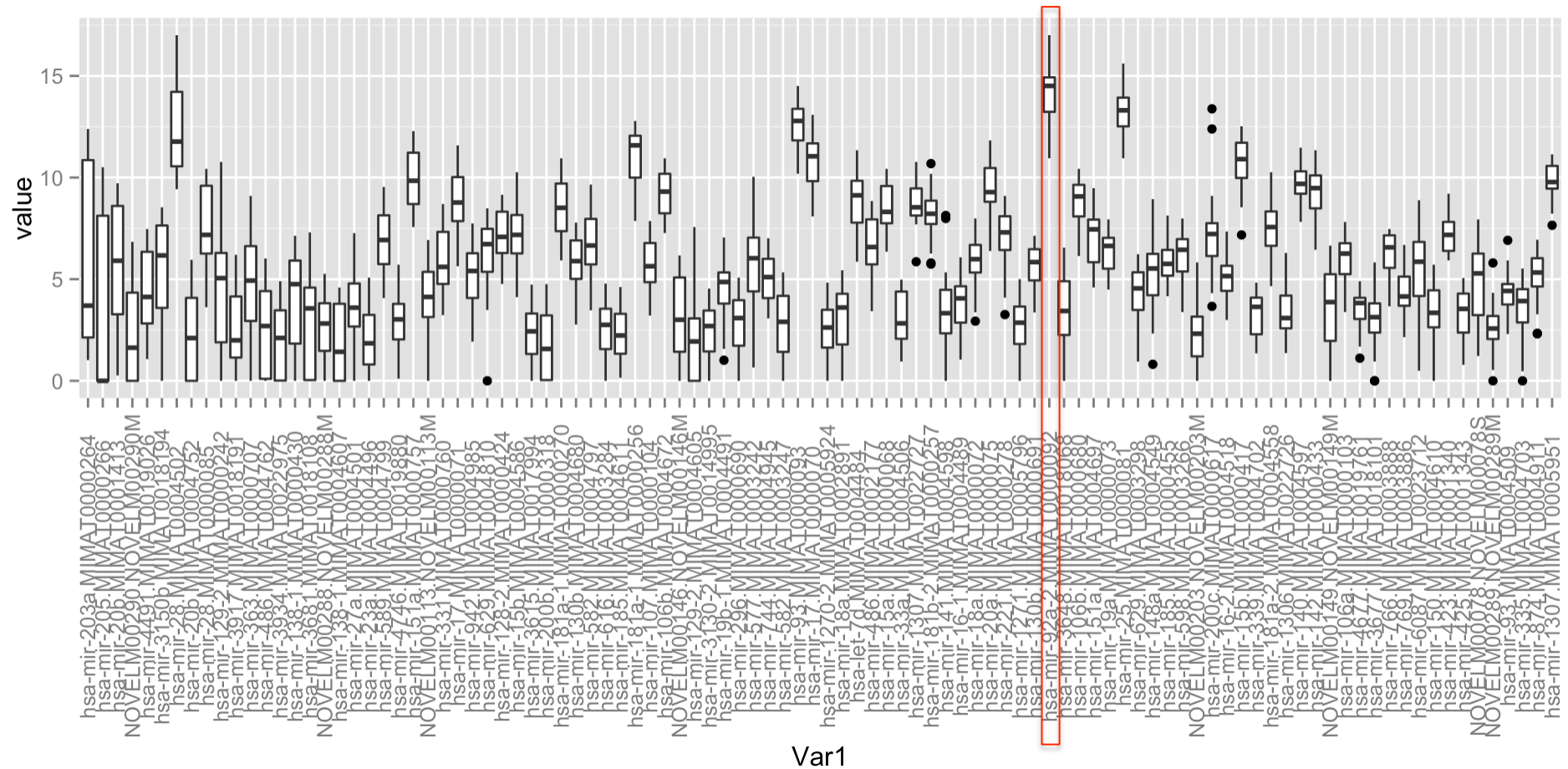
## Features

## Samples

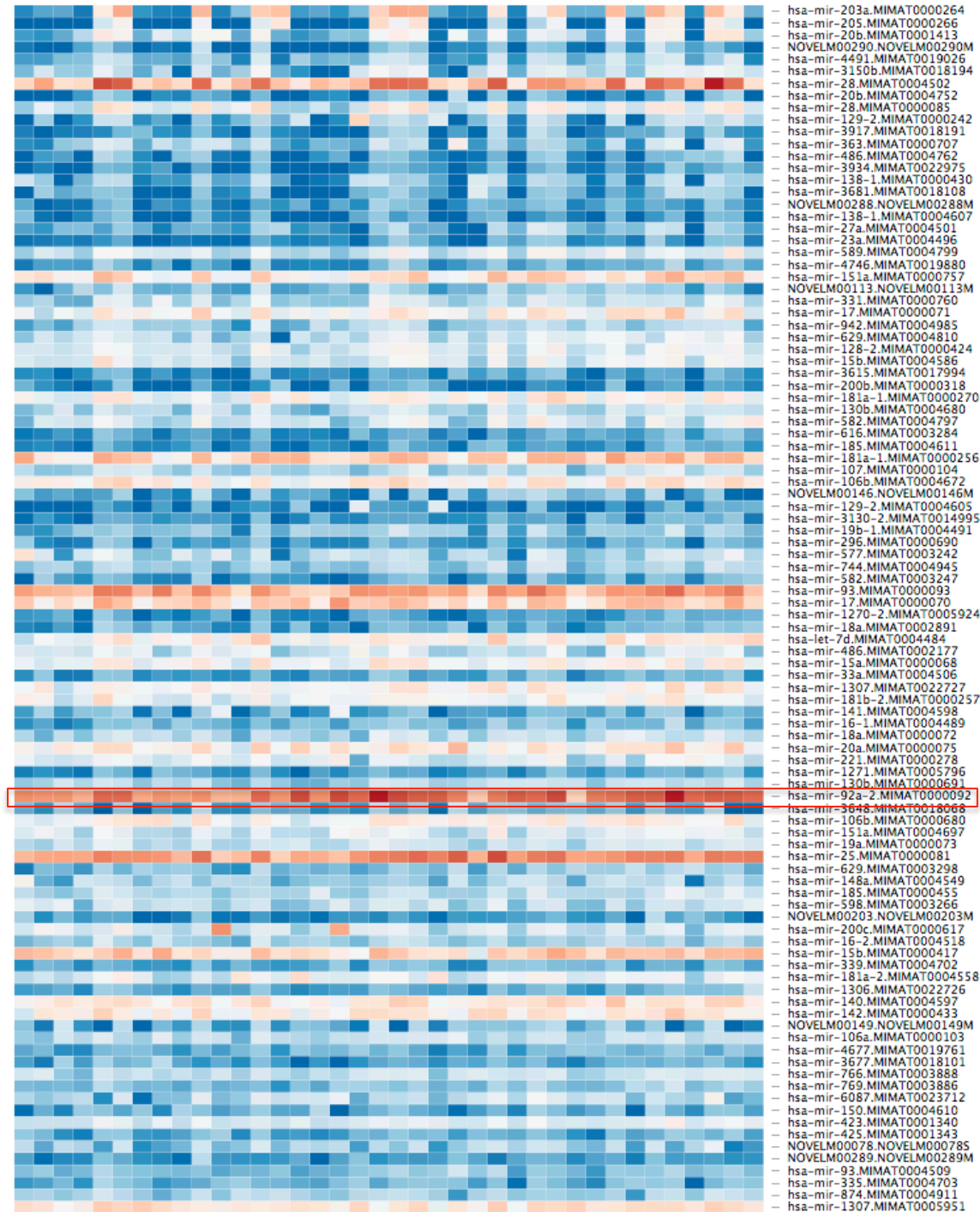
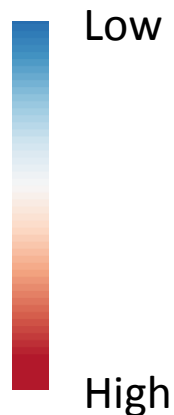
# Coding Exercise #2

- 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



What  
happened  
here?



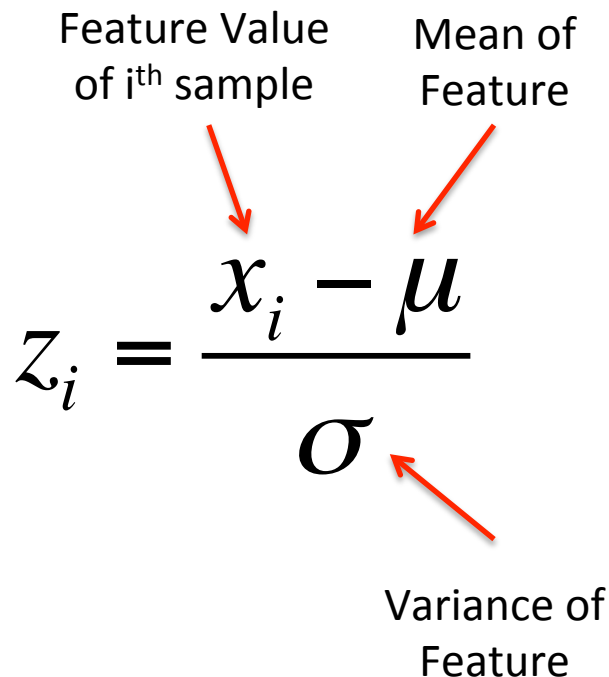


# Importance of Normalization

- 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



The diagram illustrates the Z-score normalization formula with three labels and red arrows pointing to the corresponding parts of the equation:

- Feature Value of  $i^{\text{th}}$  sample**: Points to  $x_i$  in the numerator.
- Mean of Feature**: Points to  $\mu$  in the numerator.
- Variance of Feature**: Points to  $\sigma$  in the denominator.

$$z_i = \frac{x_i - \mu}{\sigma}$$



# Bonus Question(s) / Food for Thought

- How could you perform a different type of normalization?
- Should you normalize the columns too?



# 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