

Heatmaps in R

Kayvan Jalali & Carlos Puerta

9/20/2023

1 What is a heatmap?

2 Pheatmap Package

3 Applying Heatmaps

4 Wrapping up

Section 1

What is a heatmap?

What is a heatmap?

A heatmap is graphic that will display your data in a colorful grid. This is great for seeing trends and patterns in your data.

Generally, heatmaps are used to represent data where you have 2 categorical variables with a third continuous variable, though this is not explicitly necessary.

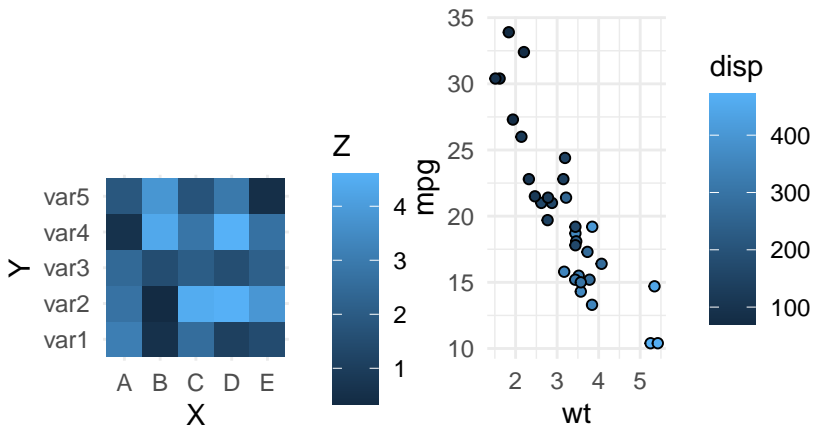
Advantages

- Adds another variable
- Easier to interpret
- Color is easy to parse

Disadvantages

- Need right choice of colors and data
- Only useful if clear
- A different plot may be better

Examples



When to use a heatmap

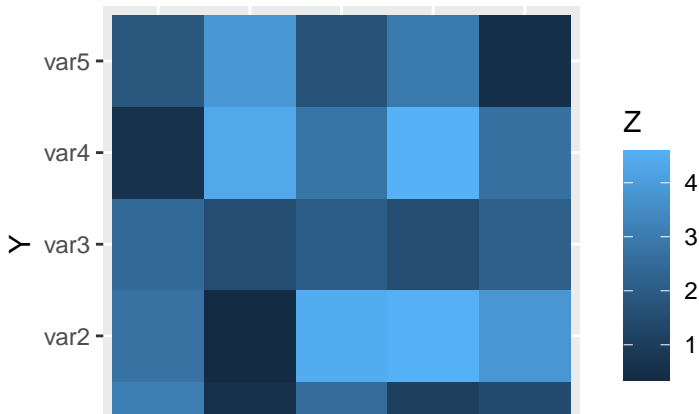
Heatmaps are applicable to both continuous and discrete variables. When working with discrete variables, distinguishing between similar colors can be challenging. To improve visibility, it is recommended to use a color palette of no more than nine colors, such as ROYGBIV, black, and white.

In the case of continuous variables, you have the option to use a sequential color scale, transitioning from a lighter hue to a darker one or vice versa. Alternatively, you can create a divergent color scale, transitioning between two distinct colors. A divergent color scale may be more useful when the center and ends of the value range are meaning full.

ggplot2

You can create a simple and quick heatmap using the `geom_tile()` function in `ggplot2`, but these heatmaps are very limited in their functionality.

```
ggplot(data, aes(X, Y, fill = Z)) +  
  geom_tile() + coord_equal()
```



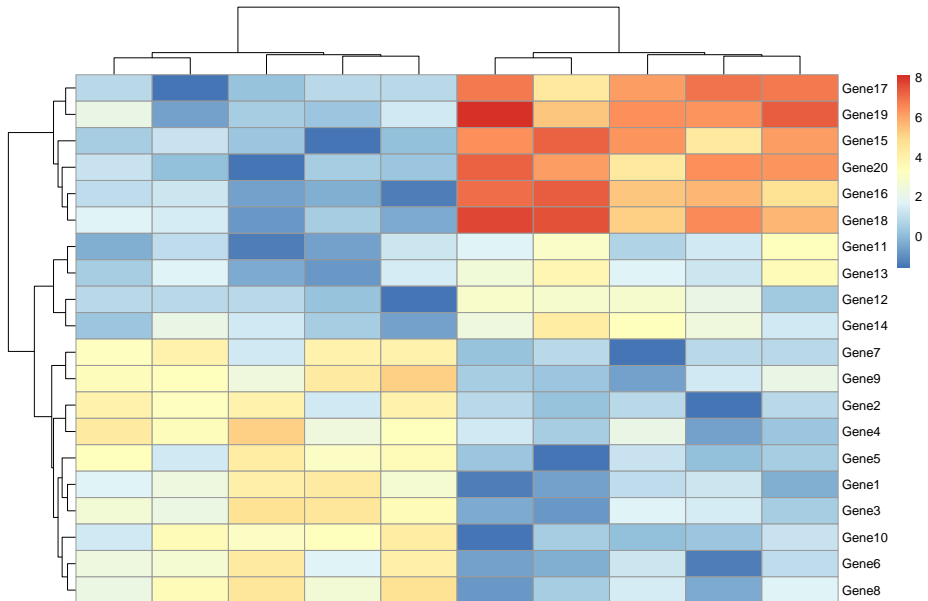
Section 2

Pheatmap Package

Pheatmap Package

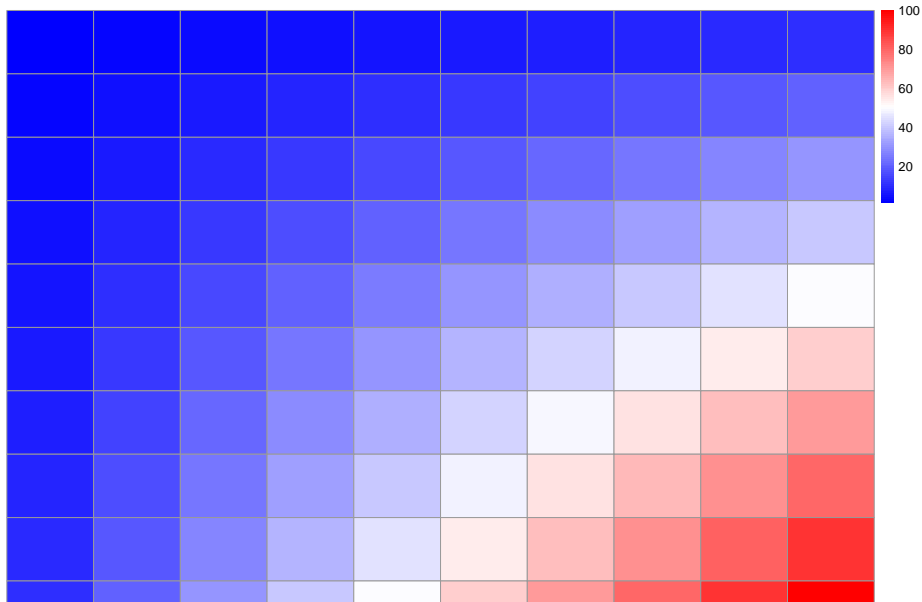
Pheatmap (Pretty Heatmaps) is a package for R that supercharges heatmaps and allows you to create incredibly complex and helpful heatmaps.

Basic functions



Dendrogram

Heatmap example

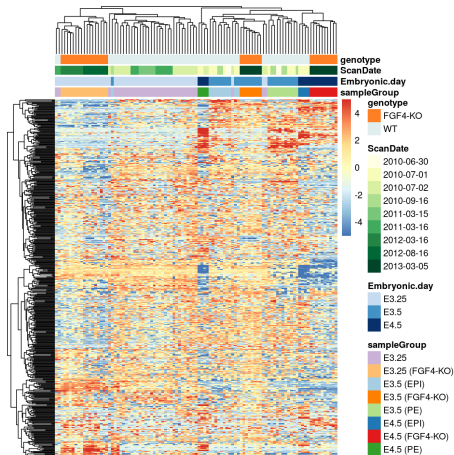


Pheatmap example



More Complex Example

Genes



Code for Genes

```

topGenes = order(rowVars(Biobase::exprs(x)),
                  decreasing = TRUE)[1:500]
rowCenter = function(x) { x - rowMeans(x) }

pheatmap(rowCenter(dfx[topGenes, ]),
  show_rownames = FALSE,
  show_colnames = FALSE,
  breaks = seq(-5, +5, length = 101),
  annotation_col = pData(x)[, c("sampleGroup",
                                "Embryonic.day",
                                "ScanDate",
                                "genotype") ],
  annotation_colors = list(
    sampleGroup = groupColor,
    genotype = c(`FGF4-KO` = "chocolate1", `WT` = "azure2"),
    Embryonic.day = setNames(brewer.pal(9, "Blues")[c(3, 6, 9)],
                             c("E3.25", "E3.5", "E4.5")),
    ScanDate = setNames(brewer.pal(nlevels(x$ScanDate), "YlGn"),

```


Section 3

Applying Heatmaps

NYC13Flights

First we get our data ready.

```
x <- flights %>%  
  filter(arr_delay > 0) %>%  
  group_by(month, carrier) %>%  
  summarize(total_delay = sum(arr_delay / 60)) %>%  
  arrange(month, carrier)
```

NYC13Flights

Then we transform it into a matrix for Pheatmap.

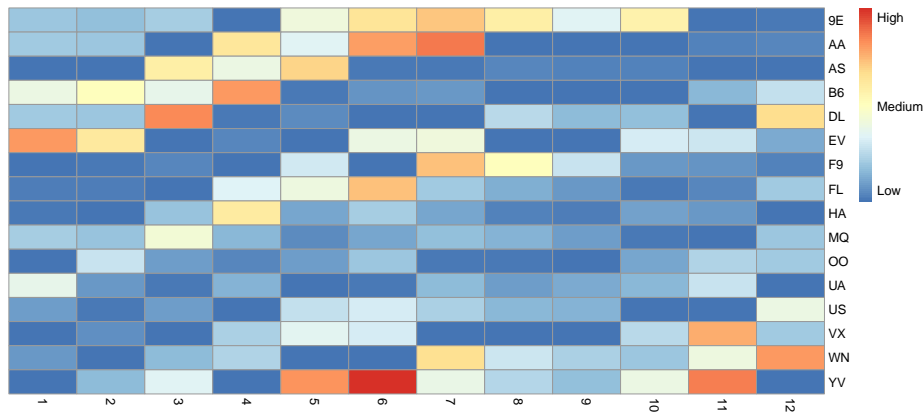
```
mat <- matrix(x$total_delay, nrow = 16)
rownames(mat) <- unique(x$carrier)
colnames(mat) <- unique(x$month)
```

NYC13Flights

And finally we plot it.

```
pheatmap(mat,  
  cluster_row = FALSE,  
  cluster_cols = FALSE,  
  legend_breaks = c(150, 1200, 2300),  
  legend_labels = c("Low", "Medium", "High")  
)
```

NYC13Flights



MPG

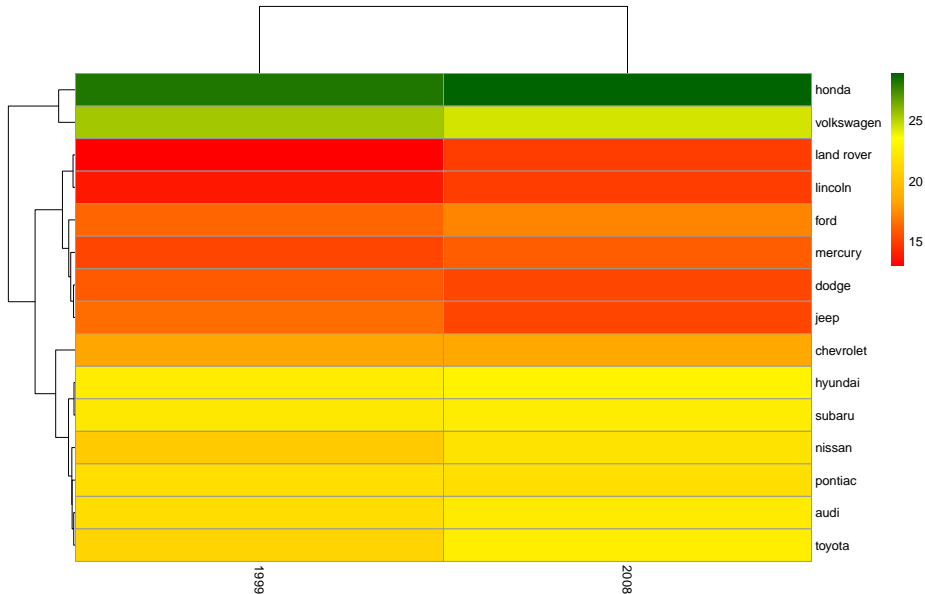
Let's do it with a simpler dataset we're familiar with. We start by getting summarizing our data.

```
x <- mpg %>%
  group_by(year, manufacturer) %>%
  mutate(avg_mpg = (cty + hwy) / 2) %>%
  summarize(avg_mpg = mean(avg_mpg)) %>%
  arrange(year, manufacturer)
```

We create a matrix and then plot it.

```
mat <- matrix(x$avg_mpg, nrow = 15)
rownames(mat) <- unique(x$manufacturer)
colnames(mat) <- unique(x$year)
pheatmap(mat,
  color = colorRampPalette(
    c("red", "orange", "yellow", "darkgreen")
  )(1000),
)
```

MPG



Section 4

Wrapping up
