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.

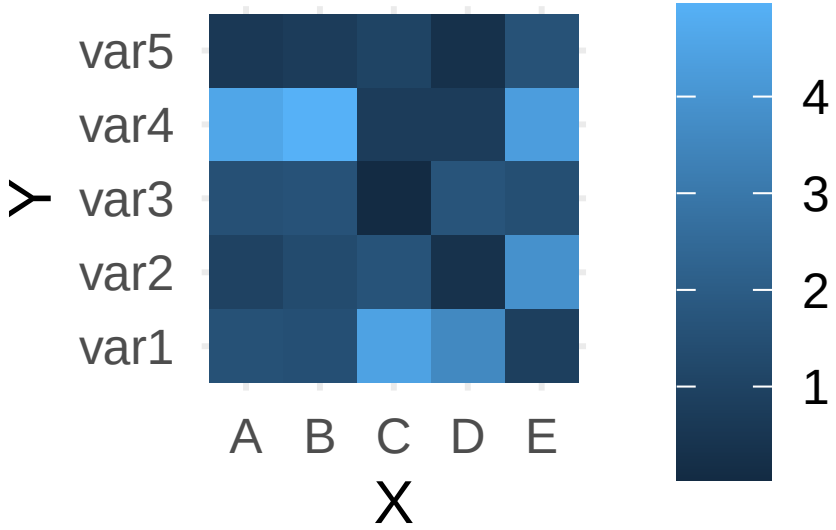
Advantages

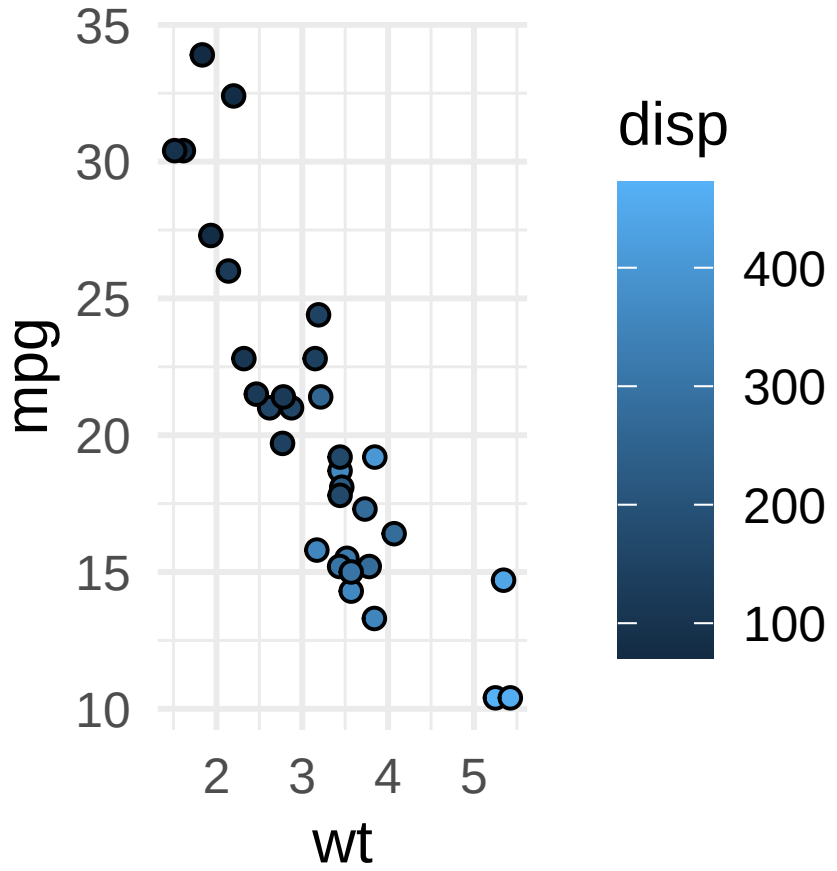
- Adv
- Adv

Disadvantages

- Disadv
- Disadv

Examples


var5
var4
var3
var2
var1
Y
A B C D E
X
Z
4
3
2
1


35
30
25
20
15
10
mpg
2 3 4 5
wt
disp
400
300
200
100

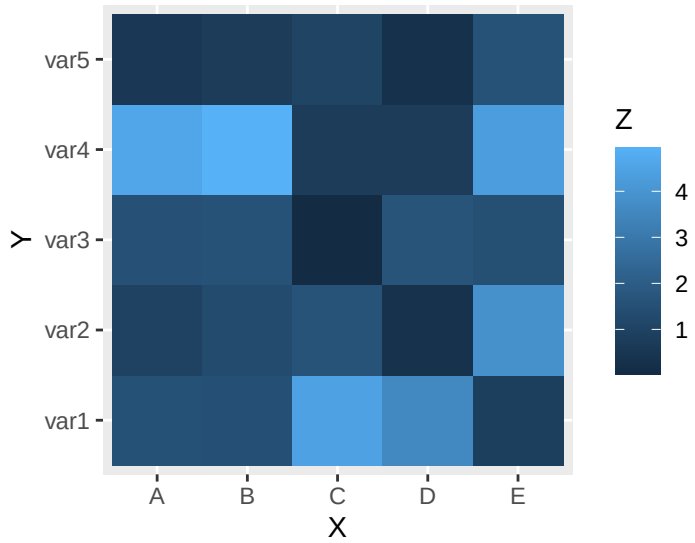
When to use a heatmap

Eventhough heatmaps can be used with continuous variables, they can be difficult to digest. Ideally, heatmaps are used to visualize two categorical variables and a single continuous variable.

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()
```



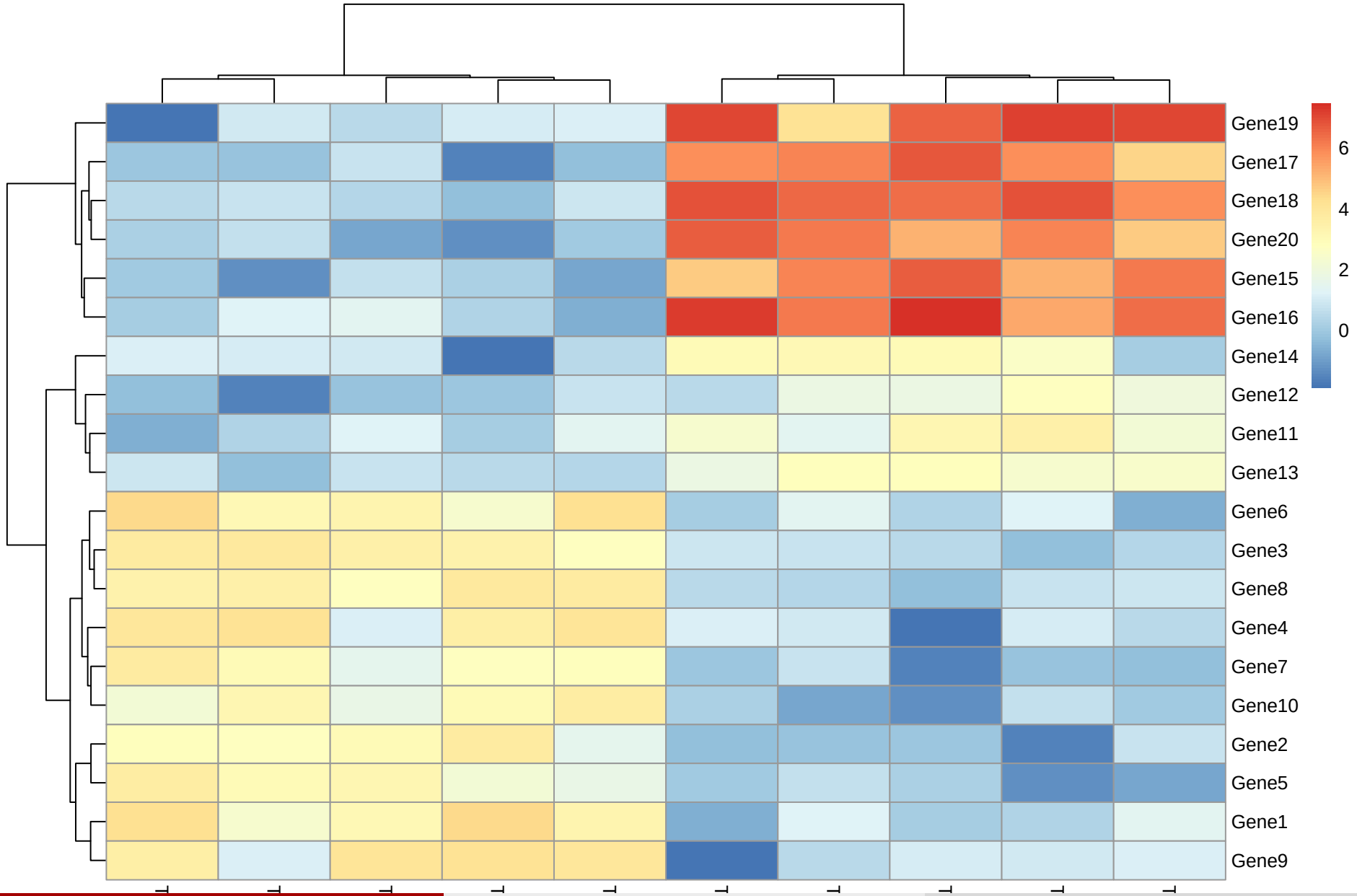
Section 2

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

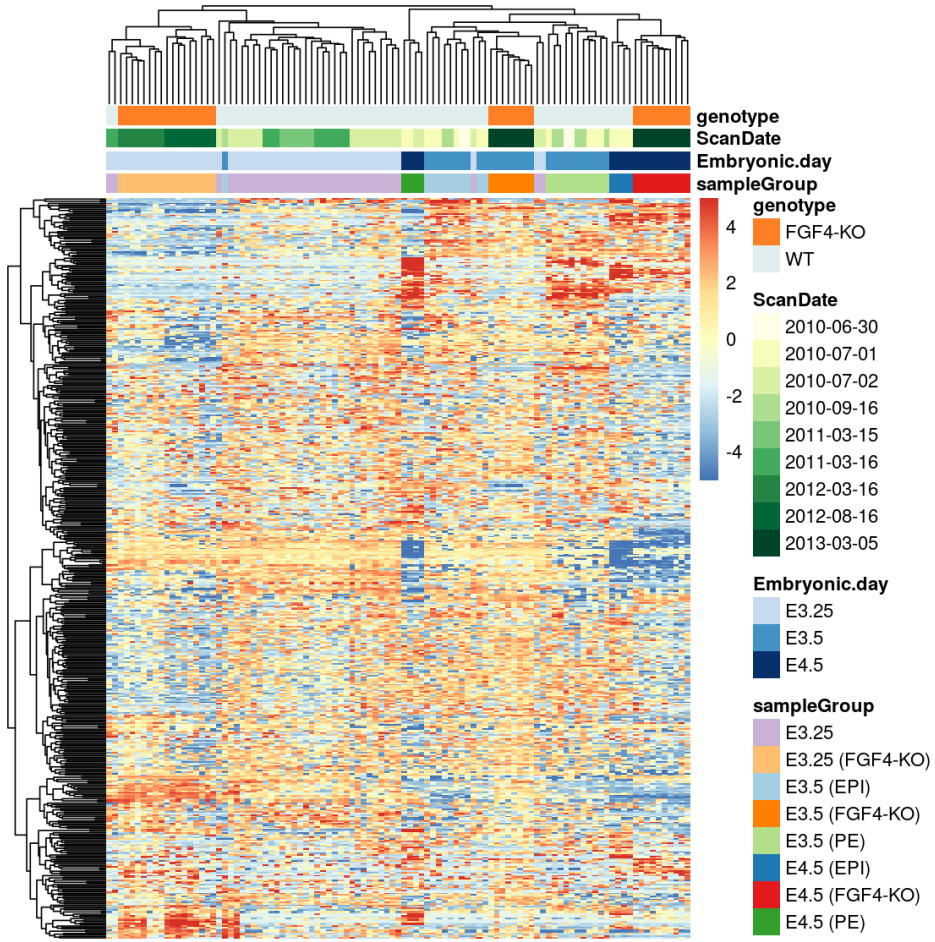


Dendogram

Example Pheatmap

More Complex Example

Genes



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.

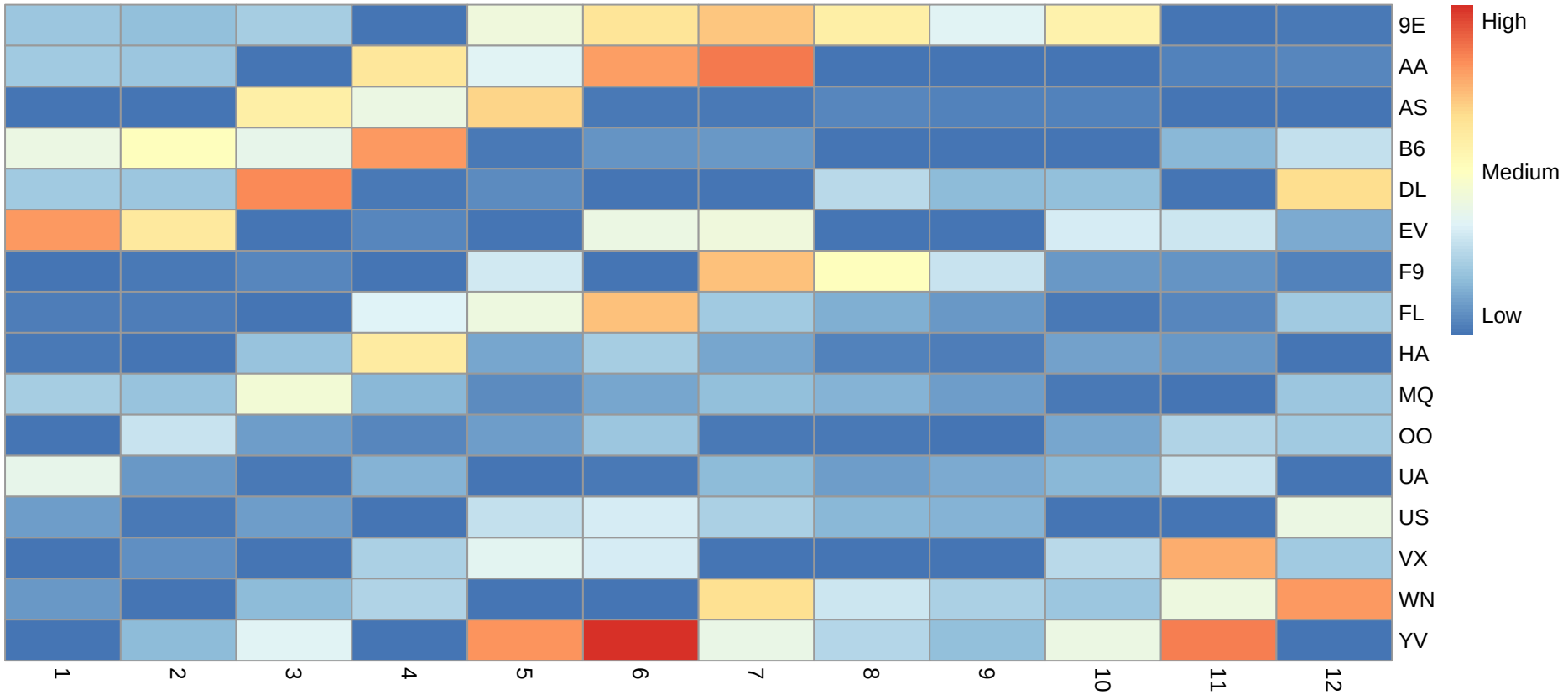
```
mat <- matrix(x$total_delay, nrow = 16)
rownames(mat) <- unique(x$carrier)
colnames(mat) <- unique(x$month)
```

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

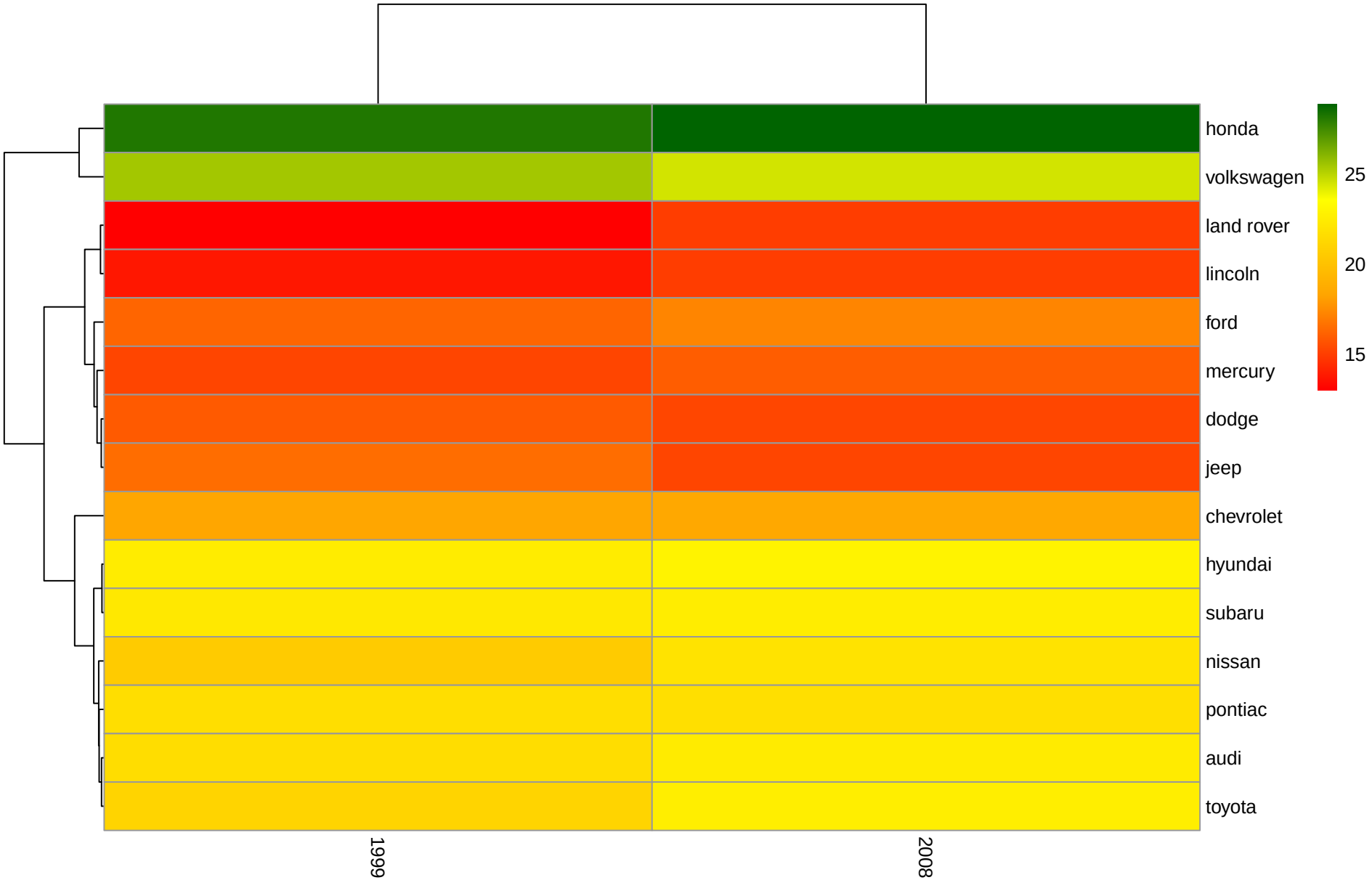
Let's do it with a simpler dataset we're familiar with. We start by getting summarizing our data.

```
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


honda
volkswagen
land rover
lincoln
ford
mercury
dodge
jeep
chevrolet
hyundai
subaru
nissan
pontiac
audi
toyota
25
20
15
1999
2008

Section 4

Wrapping up