Lab 3: Dendrograms in Heatmaps, and Matrix Multiplication

Introduction to Data Mathematics 2021

Jared Gridley

In this lab you will apply what you learned in Prelab 3. This notebook includes only **Lab** sections; you were expected to complete the separate **Prelab** work before 9 am on Wednesday. Complete the quiz by 9 am Wednesday. Questions highlighted as *Exercises* will be graded, so be sure you have answered them in your final submission! After you have completed this notebook. "Knit"" it and upload it to LMS under "Lab 3"" assignment. **Make sure that your name is at the top!** ** To get full credit on any plots, you must give them meaningful titles **

First, read in the `mtcars` dataset, which is built-in to R. Add code to scale the data and save as a matrix in mt.matrix

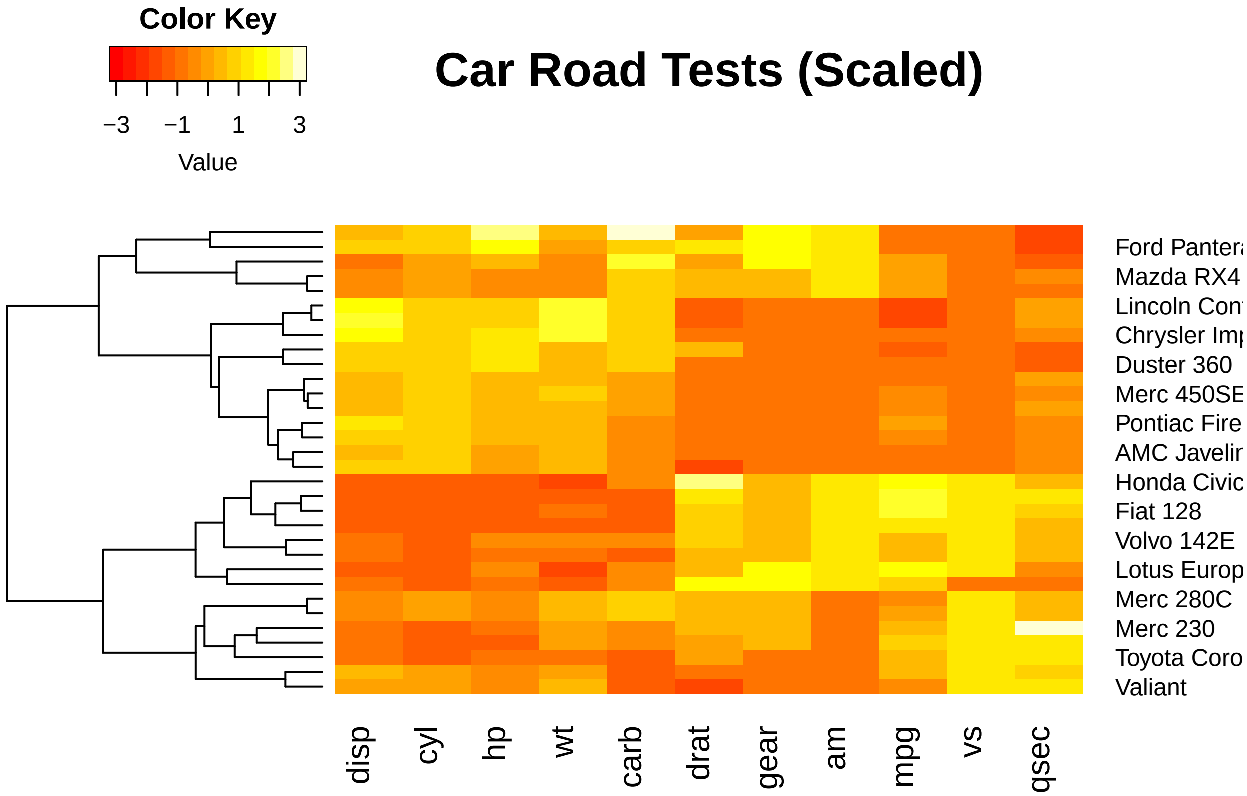
```
# Insert your code here
raw.df<-mtcars

mt.matrix <- scale(raw.df)
```

Read the documentation about the dataset so you can interpret the features, for example am = 0 means the car has an automatic transmission and am = 1 means a manual transmission: https://www.rdocumentation.org/packages/datasets/versions/3.5.2/topics/mtcars

Exercise 1

Create a heatmap of the *scaled* data with clustering of the rows and columns. Show the dendogram of the rows but not the columns. Do not use any of the built-in scaling functions in heatmap.2.

```
# Note that the heatmap in the pdf does not show all row names, but when opened
# in a seperate browser all the names are revealed in the rows.
heatmap.2(mt.matrix,
          main = "Car Road Tests (Scaled)",
          dendrogram = "row",
          tracecol = NA,
          density.info = 'none')
```



Answer the following questions regarding the heatmap with column scaling, supposing that we cut off the dendogram at 4 clusters. Put your answer after each question.

- Describe the cluster containing the Toyota Corona? State how many cars are in it and describe in words the characteristics that define this cluster.

There are 7 cars in the cluster containing the Toyota Corona. This cluster of cars has a straight shaped engine, automatic transmission, relatively average gross horsepower. This cluster of cars was not very “peppy”, that is their average 1/4 mile time was relatively high (so slower than the other clusters), which agrees with the cluster having slightly fewer than average number of cylinders. The cars in this cluster were also slghtly above average when it came to fuel efficiency (mpg).

- Describe the cluster containing the Ferrari Dino? State how many cars are in it and describe in words the characteristics that define this cluster.

The cluster containing the Ferrari Dino has 5 cars in it. These cars are fast, as seen by their low qsec value, meaning they can travel a 1/4 mile faster than the other cars on average. The cars all had a v shaped engine and slightly below average fuel efficiency (mpg). They were all manual transmission, with an above avarage number of gears, and slightly above average rear axle ratio. They were all around the average weight for cars sampled, however they had a higher gross horsepower than other cars. This follows with their above average number of cylinders. Their displacement was average but varied between cars.

- Describe the cluster containing the Merc 450SL? State how many cars are in it and describe in words the characteristics that define this cluster.

There are 12 cars in this cluster. The cars in the clusters had higher, displacement, number of cylinders, horsepower, and weight when compared to the other cars sampled. They were only slightly above average when it came to their 1/4 mile time. The cluster was below average when it came to their rear axle ratio, number of gears, a miles per gallon. There were all automatic tranmission and all had a v shaped engine.

Exercise 2

Given the matrices

$$A = \begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix}, \; B = \begin{bmatrix} 3 & 5 & 6 \\ 4 & 2 & 1 \end{bmatrix}$$

$$C = \begin{bmatrix} 1 \\ 4 \\ 3 \end{bmatrix}, \; D = \begin{bmatrix} 3 & 5 & 6 \end{bmatrix}$$

Find the following matrix sums, scalar products, and matrix products in R. Some of the matrix products might not make sense mathematically, in which R says that the matrices are non-conformable. If the matrices are non-conformable, say so, and then explain why (do not include code for noncomformable cases).

$$A + A^t$$

```
A <- matrix(c(1,2,3,4), nrow = 2)

A + t(A)
```

```
##      [,1] [,2]
## [1,]    2    5
## [2,]    5    8
```

$$2A + A^t$$

```
2*A + t(A)
```

```
##      [,1] [,2]
## [1,]    3    8
## [2,]    7   12
```

$$B + B^t$$

```
B <- matrix(c(3,5,6,4,2,1), nrow = 2)
# This is a nonconformable case beacase the sizes of the matrices do not match.
# In order to add 2 matrices, they must have the same number of rows and cols.
# Multiplication works with transposes theough.
```

$$AB$$

```
A%*%B
```

```
##      [,1] [,2] [,3]
## [1,]   18   18    5
## [2,]   26   28    8
```

$$BA$$

```
# This is nonconfomable because the sizes do not match up. To multiply 2 matrices,
# the number of columns in the first matrix must be the same as the number of rows
# in the second matrix. B = 2 x 3 and A = 2 x 2. 2 and 3 are not equal. When they
# are not equal there is not a way to do the individual dot products.
```

$$BC$$

```r
C <- matrix(c(1,4,3), nrow = 3)
B%*%C
```

```
##      [,1]
## [1,]   33
## [2,]   24
```

$$DB$$

```r
# This is nonconfomable because the sizes do not match up.
# The number of columns in the first matrix must be the same as the number of rows
# in the second matrix. D = 1 x 3 and B = 2 x 3. 3 and 2 are not equal so the dot
# products will not work out. It would work if B was transposed.

D <- matrix(c(3,5,6), nrow = 1)
# D%*%t(B) works
```

$$CB$$

```r
# This is nonconformable because of dimension inconsistency.
# The number of columns in the first matrix must be the same as the number of rows
# in the second matrix. C has 1 columns and B has 2 rows, so they are incompatable.
```

$$CD$$

```r
C%*%D
```

```
##      [,1] [,2] [,3]
## [1,]    3    5    6
## [2,]   12   20   24
## [3,]    9   15   18
```

$$DC$$

```r
D%*%C
```

```
##      [,1]
## [1,]   41
```

$$AC$$

```r
# This is nonconformable because of dimension inconsistency.
# The number of columns in the first matrix must be the same as the number of rows
# in the second matrix. A has 2 columns and C has 3 rows.
```

$$ABC$$

```r
# This is will work because the dimensions of the first multiplication are
# compatable with the second mulitplication, the resultant matrix of AB is 2x3
# and C is a 3x1 matrix so the result is the 2x1 matrix shown below.
A%*%B%*%C
```

```
##      [,1]
## [1,]  105
## [2,]  162
```

$$B^tB$$

```r
t(B)%*%B
```

```
##      [,1] [,2] [,3]
## [1,]   34   38   11
## [2,]   38   52   16
## [3,]   11   16    5
```

$$BB^t$$

```r
# matrixes are always compatable with their transposes, though they are not communicable
B%*%t(B)
```

```
##      [,1] [,2]
## [1,]   49   41
## [2,]   41   42
```

Exercise 3

- Consider the command `B[,3]`.
- Is the result of class matrix? Justify your response with code in the chunk below. If it is not of class matrix, modify the code above to force it to be a matrix with one column. Then, save the original `B[,3]` as an object called `t1` and the modified result as `t2`. If it is of class matrix, then save the original `B[,3]` as an object called `t1` and then define `t2` to be `B[,3]` forced to be a matrix.

```r
# B[,3] is a row vector (So the result is class numeric)
class(B[,3])
```

```
## [1] "numeric"
```

```r
t1 <- B[,3]
t2 <- matrix(B[,3], ncol = 1)
```

Exercise 4

- Use the `%*%` operator to perform the multiplications $t1 \times B$ and $B \times t1$. Which one (if any) works? Why?
- Use the `%*%` operator to perform the multiplications $t2 \times B$ and $B \times t2$. Which one (if any) works? Why?

```r
# t1 x B will work because it has compatible dimensions,
# t1 -> 1x2, B -> 2x3
# So t1 x B will work, but not B X t1

t1 %*% B
```

```
##      [,1] [,2] [,3]
## [1,]   11   16    5
```

```
#B %*% t1

# B X t2 and t2 x B will now work because it has compatible dimensions
# B -> 2x3, t2 = 2x1
# So no matter which comes first, the dimensions will be incompatable.

#t2 %*% B
#B %*% t2
```

Exercise 5

In class we learned that the formula for mean centering a matrix is

$$\bar{M} = M - \mathbf{1}_m \left(\frac{1}{m} \mathbf{1}_m^T M \right)$$

In R, you can generate a m by 1 matrix of ones, by typing ones(m,1). This is from the Matlab package which emulates Matlab commands in R.
Mean center the matrix 'B' using this formula. Call the new matrix 'Bcenter'. Use 'colMeans' to verify that 'Bcenter' has mean 0.

```
#Alternative to t(ones(2,1)) is to just do ones(1,2)

Bcenter <- B - ones(2, 1)%*%((1/2)*t(ones(2,1))%*%B)

colMeans(Bcenter)
```

```
## [1] 0 0 0
```