

Assignment 5

Gyaban Essilfie-Bondzie

2023-04-07

```
library(tidyverse)

## — Attaching core tidyverse packages — tidyverse
2.0.0 —
## ✓ dplyr      1.1.1      ✓ readr      2.1.4
## ✓ forcats    1.0.0      ✓ stringr    1.5.0
## ✓ ggplot2     3.4.1      ✓ tibble     3.2.1
## ✓ lubridate  1.9.2      ✓ tidyr      1.3.0
## ✓ purrr       1.0.1
## — Conflicts —
tidyverse_conflicts() —
## ✗ dplyr::filter() masks stats::filter()
## ✗ dplyr::lag()     masks stats::lag()
## ⓘ Use the conflicted package (<http://conflicted.r-lib.org/>) to force
all conflicts to become errors

str(mtcars)

## 'data.frame':   32 obs. of  11 variables:
## $ mpg : num  21 21 22.8 21.4 18.7 18.1 14.3 24.4 22.8 19.2 ...
## $ cyl : num   6  6  4  6  8  6  8  4  4  6 ...
## $ disp: num  160 160 108 258 360 ...
## $ hp  : num  110 110 93 110 175 105 245 62 95 123 ...
## $ drat: num   3.9 3.9 3.85 3.08 3.15 2.76 3.21 3.69 3.92 3.92 ...
## $ wt  : num   2.62 2.88 2.32 3.21 3.44 ...
## $ qsec: num   16.5 17 18.6 19.4 17 ...
## $ vs  : num    0  0  1  1  0  1  0  1  1  1 ...
## $ am  : num    1  1  1  0  0  0  0  0  0  0 ...
## $ gear: num    4  4  4  3  3  3  3  4  4  4 ...
## $ carb: num    4  4  1  1  2  1  4  2  2  4 ...

mtcars$car <- row.names(mtcars)
mtcars$vs <- as.factor(mtcars$vs)
str(mtcars)

## 'data.frame':   32 obs. of  12 variables:
## $ mpg : num  21 21 22.8 21.4 18.7 18.1 14.3 24.4 22.8 19.2 ...
## $ cyl : num   6  6  4  6  8  6  8  4  4  6 ...
## $ disp: num  160 160 108 258 360 ...
## $ hp  : num  110 110 93 110 175 105 245 62 95 123 ...
## $ drat: num   3.9 3.9 3.85 3.08 3.15 2.76 3.21 3.69 3.92 3.92 ...
## $ wt  : num   2.62 2.88 2.32 3.21 3.44 ...
```

```
## $ qsec: num 16.5 17 18.6 19.4 17 ...
## $ vs : Factor w/ 2 levels "0","1": 1 1 2 2 1 2 1 2 2 2 ...
## $ am : num 1 1 1 0 0 0 0 0 0 0 ...
## $ gear: num 4 4 4 3 3 3 3 4 4 4 ...
## $ carb: num 4 4 1 1 2 1 4 2 2 4 ...
## $ car : chr "Mazda RX4" "Mazda RX4 Wag" "Datsun 710" "Hornet 4 Drive"
...
```

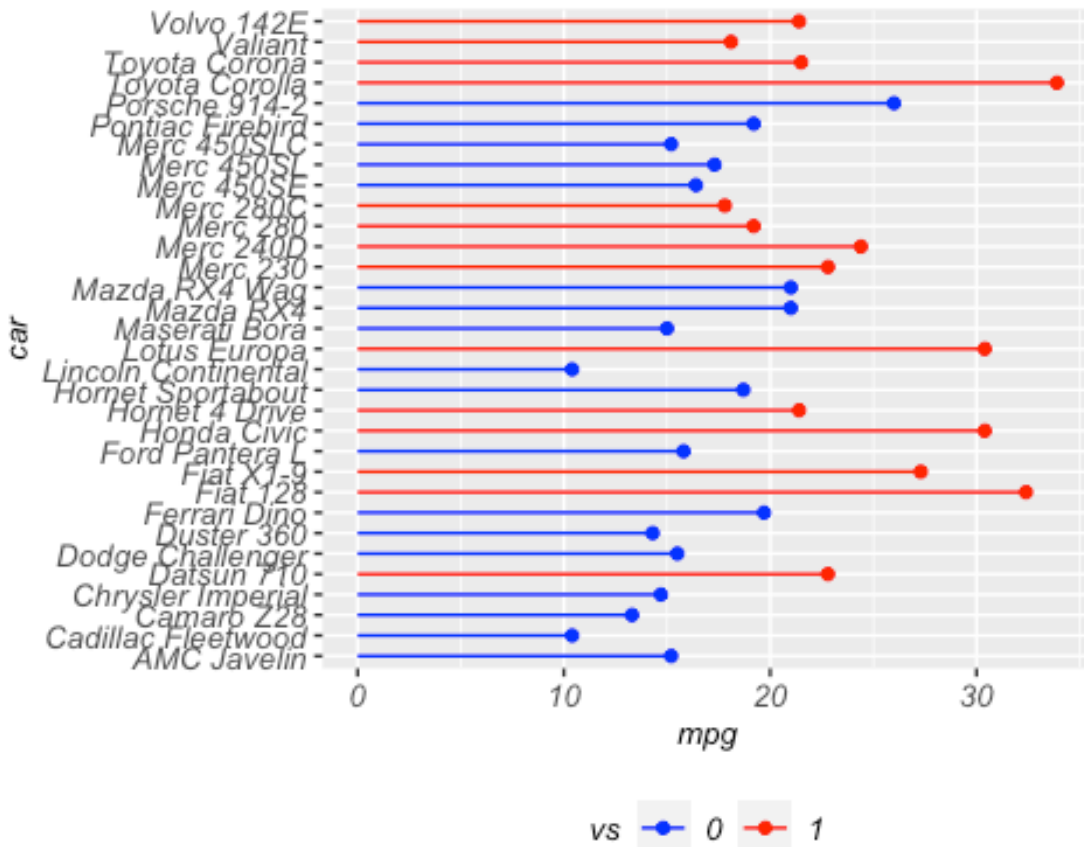
#Question 1

#3 This line copies the row names in the dataset mtcars and places them in a new column called car.

#4 The as.factor function converts the column vs into a factor or a categorical variable and the command mtcars\$vs <- returns it which makes this change permanent.

#5 We now have 12 columns due to the column car being added to the dataset and the vs column is no longer num variable, but it is a factor.

```
ggplot(mtcars, aes(x = mpg, y = car, color = vs)) + geom_segment(aes(x = 0, y
= car, xend = mpg, yend = car)) + geom_point() + scale_colour_manual(values =
c("blue", "red")) + theme(axis.title = element_text(face="italic", size=10),
axis.text = element_text(face="italic", size=10), legend.title =
element_text(face="italic", size=10), legend.text=element_text(face="italic",
size=10), legend.position="bottom")
```



#7

Theme layer added to the plot in the command above

```
library(pheatmap)
?pheatmap
gene_exp <- read.table(file =
"Desktop/Biomedical_Genomics/gene_expression.txt", header=T,
row.names="gene")
gene_exp_sub <- as.matrix(gene_exp[rowSums(gene_exp)>50000,])
my_sample_col <- data.frame(sample = rep(c("tumour", "normal"), c(4,2)))
my_sample_col

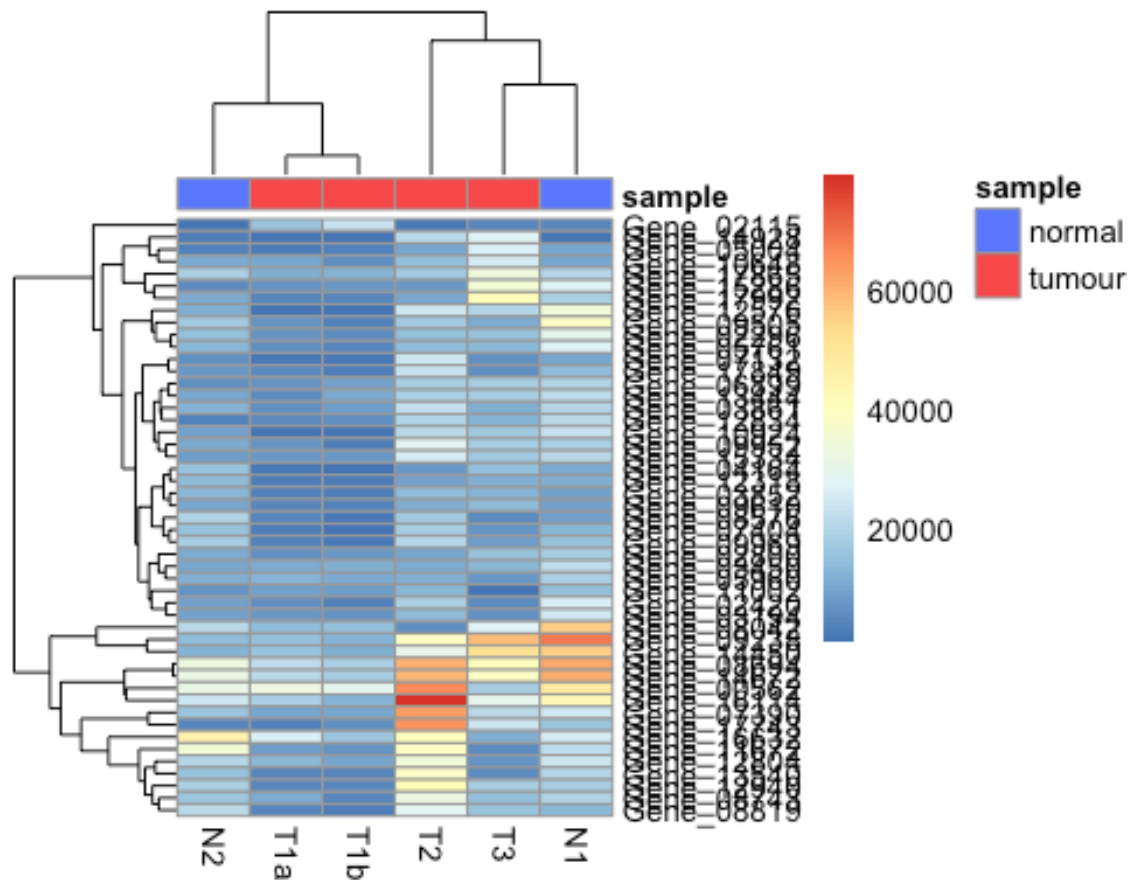
## sample
## 1 tumour
## 2 tumour
## 3 tumour
## 4 tumour
## 5 normal
## 6 normal

row.names(my_sample_col) <- colnames(gene_exp_sub)
my_sample_col

## sample
## T1a tumour
```

```
## T1b tumour
## T2 tumour
## T3 tumour
## N1 normal
## N2 normal

my_colour = list(sample = c(normal = "#5977ff", tumour = "#f74747"))
pheatmap(gene_exp_sub, annotation_col = my_sample_col, annotation_colors =
my_colour) #final plot
```



#Question 2

#6 This line allows us to read the txt file into a data frame called gene_exp.

#7 This line returns each row with a sum > 50000 from the data frame gene_exp into a matrix called gene_exp_sub.

#8 This line creates a data frame with 1 column called sample and 6 rows. The column sample has the words tumour repeated 4 times and the word normal repeated twice.

#10 The column names of the gene_exp_sub matrix become the row names of the my_sample_col data frame.

#12 It sets the different colors which will represent the separate values of the sample column in the plot. Tumour represented by the color red and normal with the color blue.