

# Other Visualization Methods

Prepared by Dr. Jose Isagani Janairo

2023

This document is for the exclusive use of students enrolled in the course Data Science for Life Scientists at De La Salle University.

So far, we have learned how to create the common types of plots for data visualization using base R and ggplot2. In this lesson, we will learn how to create **heat maps** and **chord diagrams** which are commonly used in life science research.

## Heat maps

Heat maps rely on variations on color intensity to denote differences in the values being compared and reported. Since it relies on color for visual comparison, it can be effective for the quick comparison of data among observations.

To get started with creating heat maps, let us load the needed packages for this activity.

```
library(gplots)
```

```
##
## Attaching package: 'gplots'
```

```
## The following object is masked from 'package:stats':
##
##     lowess
```

```
library(RColorBrewer)
```

The package gplots enables us to quickly create a heatmap while the RColorBrewer is a package that lets us use different color palettes for our plots. Considering how important color is for heat map, having more choices is definitely better. After loading the packages, let us load and explore the dataset.

```
univ <- read.csv(file = "university.csv", header = TRUE)
univ
```

```
##   University LifeScience PhysicalScience Health   IT Engineering SocialScience
## 1          A         0.745           0.632 0.432 0.633      0.345      0.275
## 2          B         0.825           0.701 0.532 0.263      0.447      0.952
## 3          C         0.634           0.846 0.789 0.189      0.424      0.362
## 4          D         0.846           0.981 0.789 0.792      0.789      0.370
## 5          E         0.231           0.858 0.767 0.870      0.309      0.109
## 6          F         0.567           0.721 0.845 0.983      0.861      0.344
## 7          G         0.341           0.821 0.678 0.224      0.198      0.689
## Humanities Business
## 1         0.225      0.824
## 2         0.132      0.539
## 3         0.762      0.814
## 4         0.782      0.675
## 5         0.835      0.885
## 6         0.112      0.900
## 7         0.232      0.539
```

This is a hypothetical dataset that compares the ratings of Universities A to G on different subjects. The ratings are numerical values ranging from 0 to 1. To create the heat map, we prepare the dataset by executing the following command:

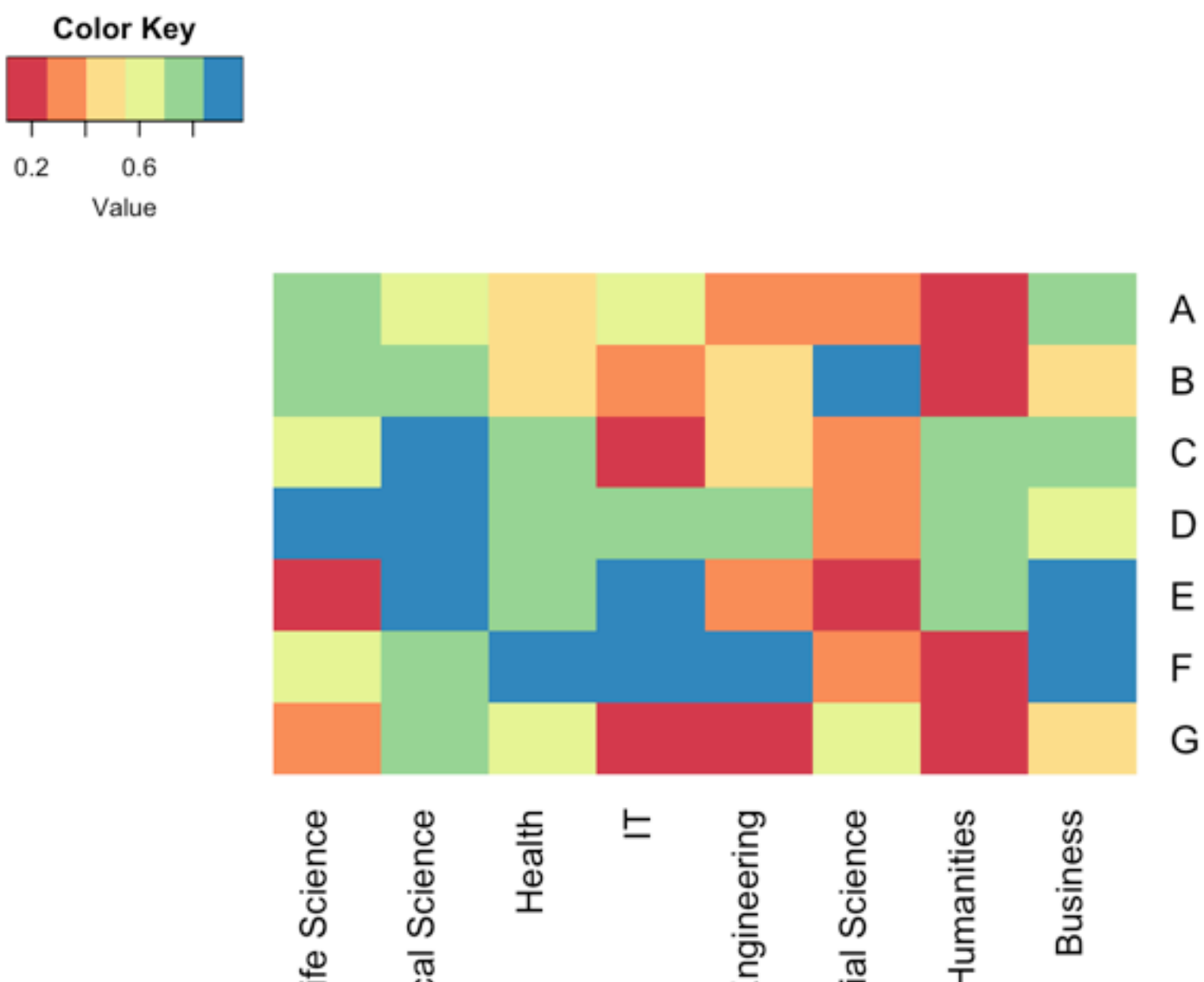
```
row.names(univ) <- univ$University
univ[1] <- NULL
univ1 <- as.matrix(univ)
subjects <- c("Life Science", "Physical Science", "Health", "IT", "Engineering", "Social Science", "Humanities", "Business")
```

**Hands-on Output 1: Why was the previous set of commands executed?**

```
heatmap.2(univ1,col=brewer.pal(6,"Spectral"), trace = "none", Colv = NA, Rowv = NA, density.info = "none", margin
s=c(7,5), labCol = subjects)
```

```
## Warning in heatmap.2(univ1, col = brewer.pal(6, "Spectral"), trace = "none", :
## Discrepancy: Rowv is FALSE, while dendrogram is `both`. Omitting row dendrogram.
```

```
## Warning in heatmap.2(univ1, col = brewer.pal(6, "Spectral"), trace = "none", :
## Discrepancy: Colv is FALSE, while dendrogram is `column`. Omitting column
## dendrogram.
```

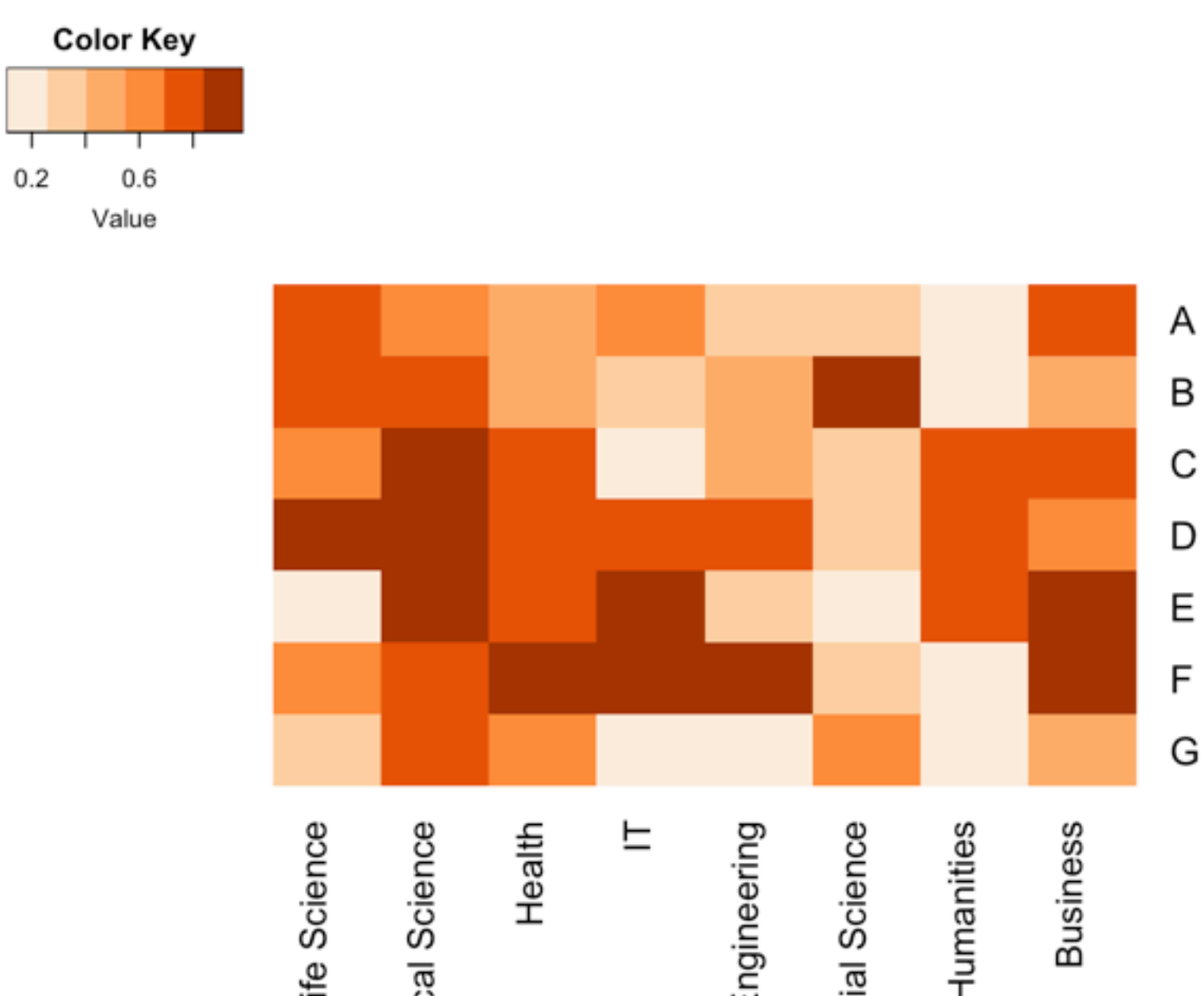


The heat map is visually appealing, but it's too colorful in my opinion. The goal of heat maps is to quickly compare values, so it is advisable to use color schemes that will be able to achieve this. Let us try various shades of orange.

```
heatmap.2(univ1,col=brewer.pal(6,"Oranges"), trace = "none", Colv = NA, Rowv = NA, density.info = "none", margins
=c(7,5), labCol = subjects)
```

```
## Warning in heatmap.2(univ1, col = brewer.pal(6, "Oranges"), trace = "none", :
## Discrepancy: Rowv is FALSE, while dendrogram is `both`. Omitting row dendrogram.
```

```
## Warning in heatmap.2(univ1, col = brewer.pal(6, "Oranges"), trace = "none", :
## Discrepancy: Colv is FALSE, while dendrogram is `column`. Omitting column
## dendrogram.
```



Do you think this is better in terms of ease in comparing the different universities?

**H002: Recreate the heat map using a different color palette. Consult RColorBrewer for the different palettes you can use.**

## Chord diagram

For this part of the activity, we will recreate a chord diagram from our recent publication. The full text of this paper has been provided to you. Skim through the paper and see how the chord diagram was used.

**H003: How did the paper use chord diagrams? What information or data was summarized/visualized through this type of plot?**

To get started in creating chord diagrams, we will use the package circize and load the dataset. Make sure you understood how chord diagrams were used in the paper before proceeding.

```
library(circize)
```

```
## Warning: package 'circize' was built under R version 4.1.2
```

```
## =====
## circize version 0.4.15
## CRAN page: https://cran.r-project.org/package=circize
## Github page: https://github.com/jokergoo/circize
## Documentation: https://jokergoo.github.io/circize_book/book/
##
## If you use it in published research, please cite:
## Gu, Z. circize implements and enhances circular visualization
##   in R. Bioinformatics 2014.
##
## This message can be suppressed by:
## suppressPackageStartupMessages(library(circize))
## =====
```

```
df <- read.csv("chorddiagnostics.csv")
head(df)
```

```
##   Sample SVM DL LASSO LDA NB Ensemble Nnet
## 1 Saliva  1 0    0 0 0    0 1
## 2 Histopath 0 9    0 0 0    0 0
## 3 Genetic  2 0    0 0 0    0 0
## 4 CT      2 1    1 0 0    0 0
## 5 Blood  0 0    0 1 0    2 0
## 6 MRI     0 0    0 0 1    0 0
```

**H004: Explain the contents of the dataset and the information they convey.**

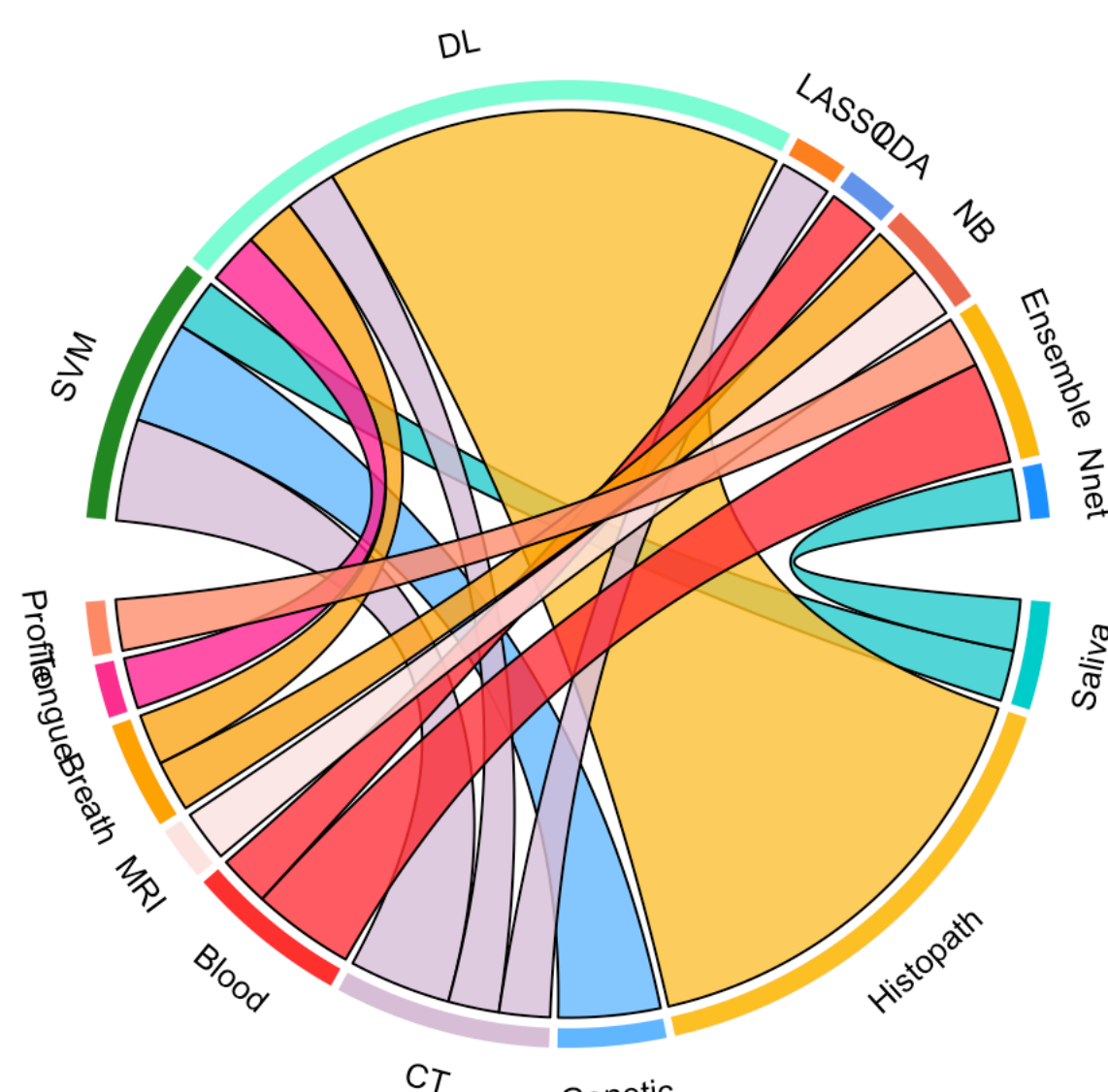
After loading the dataset, we prepare it and reformat it to the requirements.

```
row.names(df) <- df$Sample
df[1] <- NULL
df <- as.matrix(df)
```

In creating the chord diagram, we can let R decide on the default color scheme. However, it is best to be consistent in the colors we used. Look at the 3 chord diagrams used in the paper and we can see that the colors represent the same items all throughout the text. This is important to convey clarity and consistency in reporting the data.

```
colors <- c(SVM = "forestgreen", DL = "aquamarine1", LASSO = "chocolate1",
            LDA = "cornflowerblue", NB = "coral2", Ensemble = "darkgoldenrod1",
            Nnet = "dodgerblue", Saliva = "cyan3", Histopath = "goldenrod1",
            Genetic = "steelblue1", CT = "thistle", Blood = "firebrick1",
            MRI = "mistyrose1", Breath = "orange1", Tongue = "deeppink",
            Profile = "salmon1")

chordDiagram(df, annotationTrack = c("name", "grid"), grid.col = colors, transparency = 0.2,
            link.lwd = 1,
            link.lty = 1,
            link.border = 1)
```



## Exercises

- Using the university.csv dataset, create a bar plot to demonstrate that the information conveyed by heat maps can also be conveyed by bar plots. When should you use one over the other?
- Create your own heat map. You can create a hypothetical dataset or download an existing dataset from the internet (cite the source). Show the code and explain the data being visualized.
- What are the limitations of chord diagrams?
- Create your own chord diagram. You can create a hypothetical dataset or download an existing dataset from the internet (cite the source). Show the code and explain the data being visualized.

This document is for the exclusive use of students enrolled in the course Data Science for Life Scientists at De La Salle University.