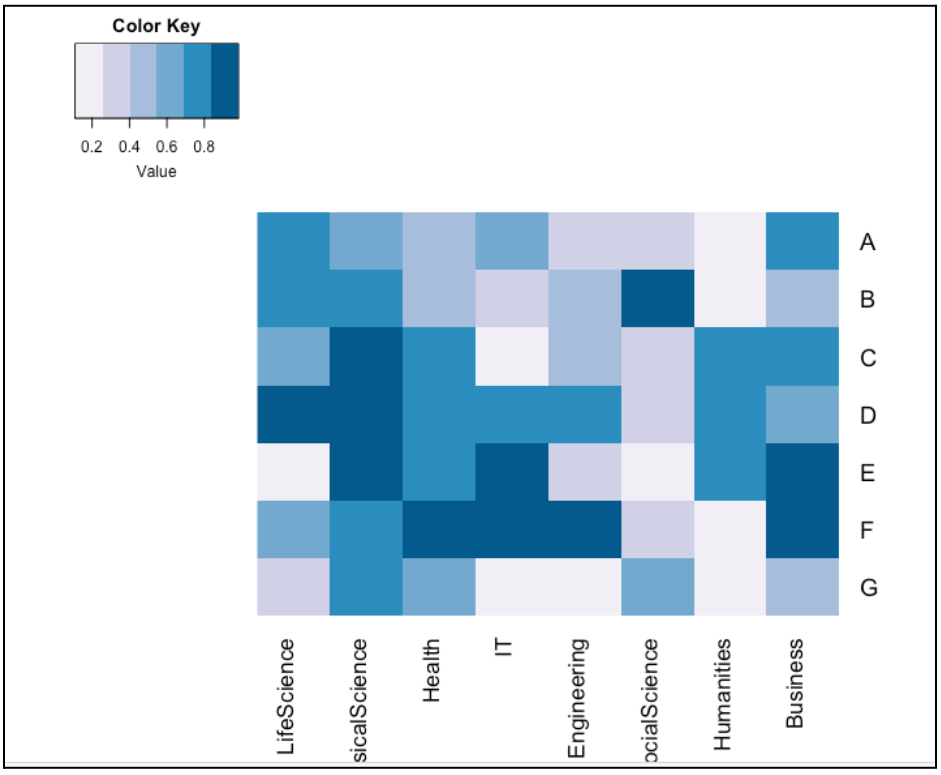
BONBUTTERY

Other Visualization Exercises

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

- The previous set of commands allows you to set up the variables and vectors needed for the creation of the heat map. The subject vector was created as the names for the x-axis and the as.matrix in order to convert the objects into matrices that can be plotted in the heatmap. Also, the row.names function changed the row names into the university names/letters (i.e. letters A to G).

HOO2: Recreate the heat map using a different color palette. Consult RColorBrewer for the different palettes you can use.



Code: **> heatmap.2(univ1,col=brewer.pal(6,"PuBu"), trace = "none", Colv = NA, Rowv = NA, density.info = "none", margins=c(7,5))**

HOO3: How did the paper use chord diagrams? What information or data was summarized/visualized through this type of plot?

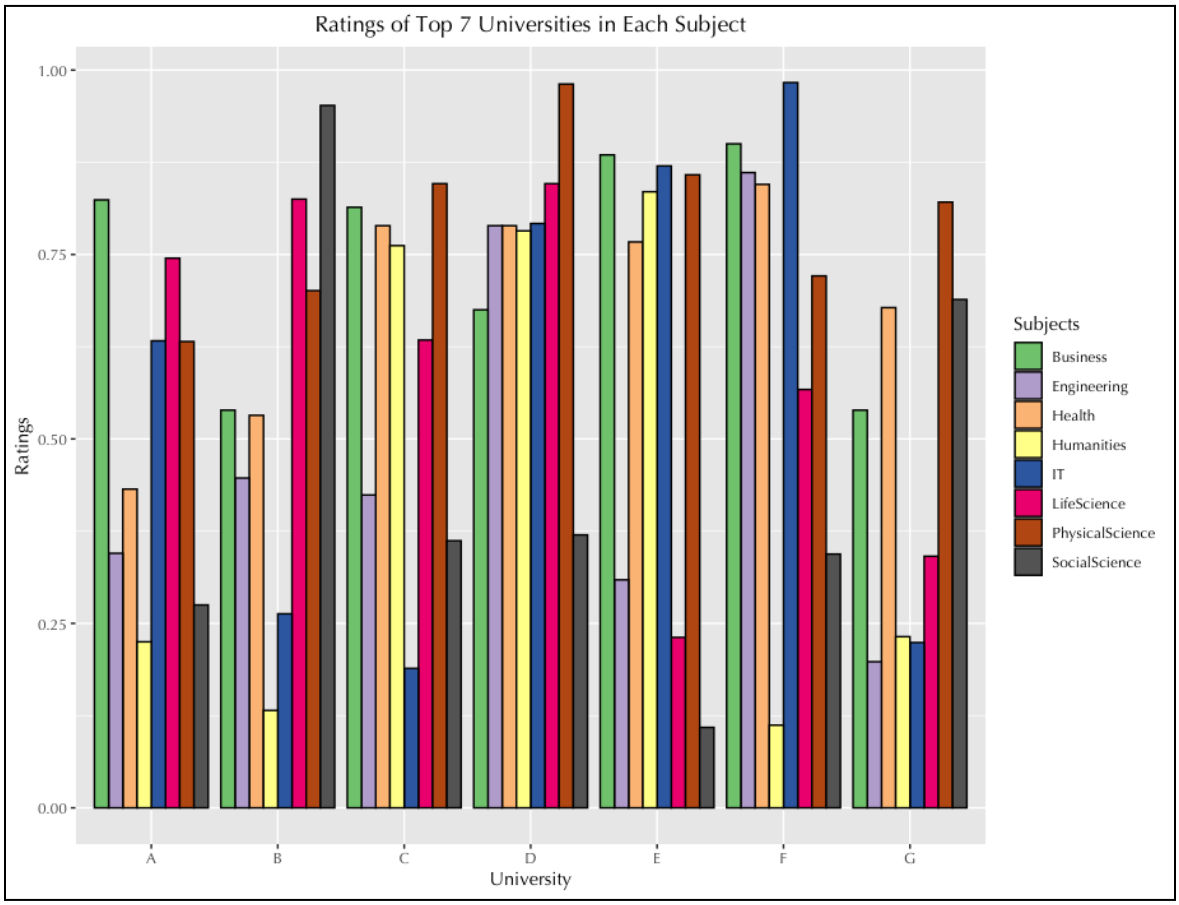
- The chord diagrams used in the paper showed the relationship between the specimen types and the algorithms used for the application of machine learning to non-endoscopically detect gastric cancer and its subtypes. Different machine learning algorithms were used, including Deep Learning (DL), Linear discriminant analysis (LDA), etc., to see in which specimens they are usually applied and explored, such as histopathological images, breath, CT scans, etc. This chord diagram allows the summarization of the flow of the data used, in this specific case, the machine learning for the different non-endoscopic gastric cancer specimens.

HOO4: Explain the contents of the dataset and the information they convey.

- The csv file dataset, “chorddiagnostics.csv” shows the different samples that are somehow similar to the ones on the paper wherein different specimen samples and machine learning (ML) algorithms are being plotted to see any possible relationships among them. Samples are categorical in nature, while each column of the algorithms is numerical (i.e. discrete) since it is looking at the number of times the algorithms are used in the detection/diagnosis of gastric cancer on the different samples.

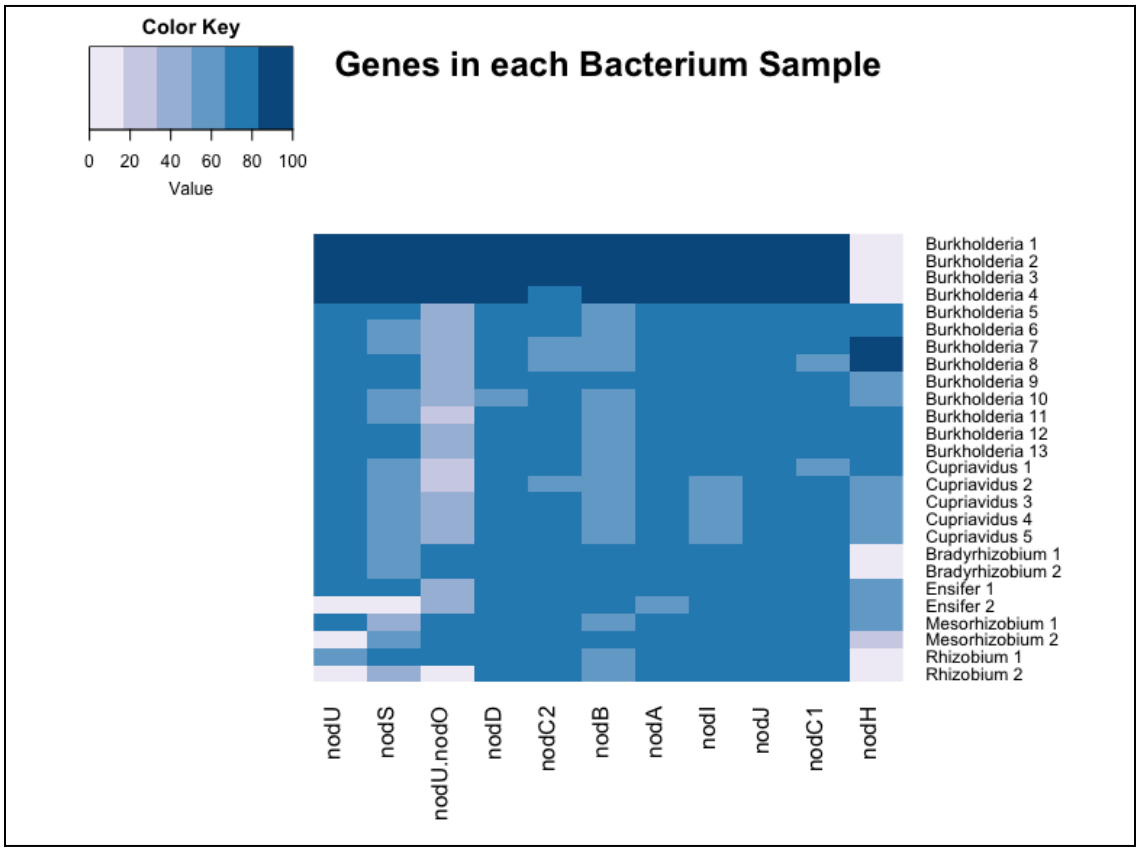
Exercises:

1. **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?**
    - Although both plots can be used to express the same kind of data, they have their own pros and cons that can help us understand when to use one over the other. According to Yi (n.d.), a grouped bar chart is a good alternative for heat maps, however, if we want a more precise comparison that can easily be visualized, a grouped bar chart is more recommended. On the other hand, it is more preferred to use heat maps when both axes are numeric in nature, since it has a better ability to visualize an overview of both axes at the same time.



Code: > ggplot(data = univ2, aes(x = University, y = Ratings, fill = Subjects))+ geom_bar(position="dodge", stat="identity", color = "Black") + theme(text = element_text(family = "Optima")) + ggtitle("Ratings of Top 7 Universities in Each Subject") + theme(plot.title = element_text(hjust = 0.5)) + scale_fill_brewer(palette="Accent")

2. **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.**
    - Data Source: Leah Briscoe (2016) from github.com
    - The heatmap below shows the percentages of each gene in different samples of bacteria.



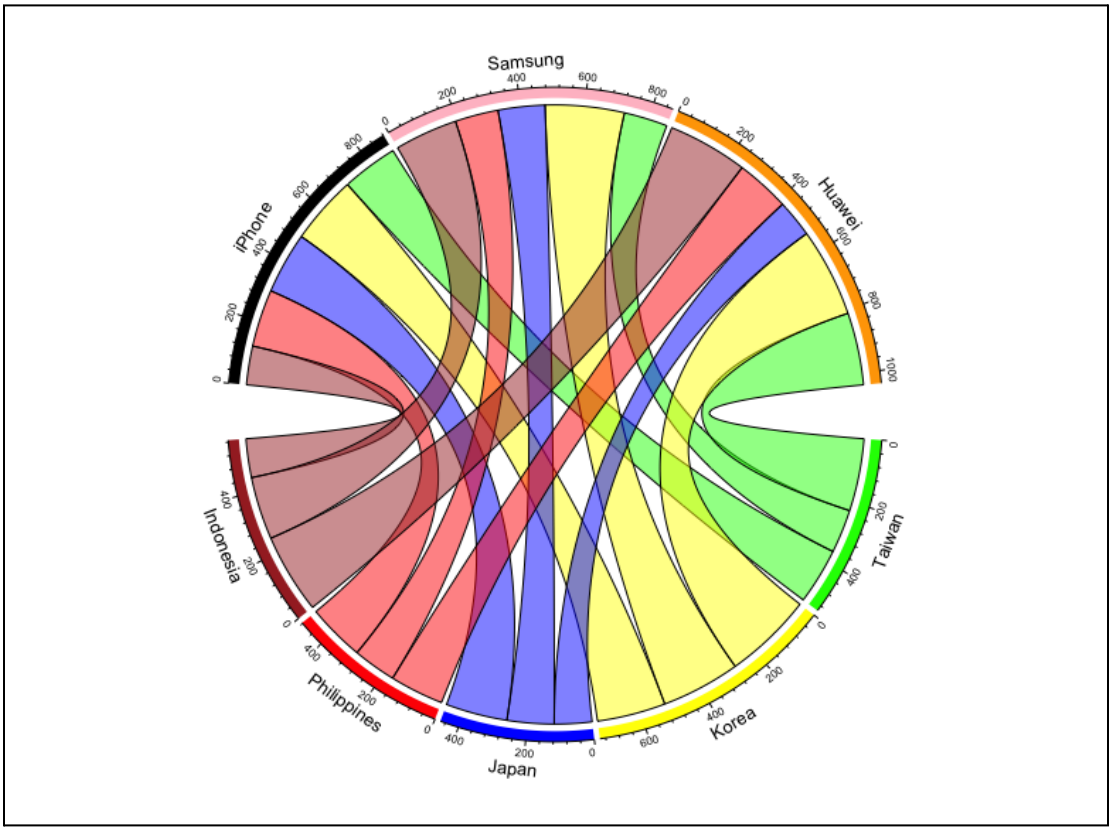
Code:

```
> Nodgenes1 <- data.matrix(NodGenes)

> heatmap.2(Nodgenes1, main = "Gene Samples", trace = "none", margins = c(7,10), Colv
= NA, Rowv = NA, density.info = "none", col=brewer.pal(6,"PuBu"))
```

3. **What are the limitations of chord diagrams?**
    - I think one of the downsides of a chord diagram is when the relationships or connections between the data become too numerous, it would be somehow difficult to understand and see correctly the relationships as it will become too cluttered. Additionally, if you want to immediately express a story or present to an audience, chord diagrams may not be a good fit for plotting as it would take a little more time to let the viewers understand.

4. **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.**
    - Data source: Created a sample data, but inspired from Rise Academy (2019)
    - The constructed sample chord diagram below shows the top 5 countries that utilize the top 3 smartphone brands (users are in thousands): Philippines, Indonesia, Japan, Korea, and Taiwan. [sample data only]



Code:

```
random_values <- c(300:100)
```

```
random_sample <- sample(random_values, 15)


col.pal = c(Taiwan = "Green", Korea = "Yellow", Japan = "Blue", Philippines = "Red",
Indonesia = "Brown",
        iPhone = "Black", Samsung = "Pink", Huawei = "Orange")


sample_matrix <- matrix(random_sample, nrow = 5,
                dimnames=list(c("Taiwan", "Korea", "Japan", "Philippines", "Indonesia"),
                        c("iPhone", "Samsung", "Huawei")))


chordDiagram(sample_matrix, grid.col = col.pal,
        link.lwd = 1,
        link.lty = 1,
        link.border = 1)
```

REFERENCES

Briscoe, L. (2016). *Sample data for "Make a Heatmap on R studio"*. Gist. https://gist.github.com/LeahBriscoe/daea6f00f78c0ccca245f8c1bae776bf#file-nodgenes-csv

The Rise Academy. (2019). *How to create Chord Diagram using R | Data Science* [Video]. YouTube. https://www.youtube.com/watch?v=_DXq8f3TNp8&t=83s&ab_channel=TheRiseAcademy

Yi. (n.d.). *A complete guide to Heatmaps*. Chartio. https://chartio.com/learn/charts/heatmap-complete-guide/