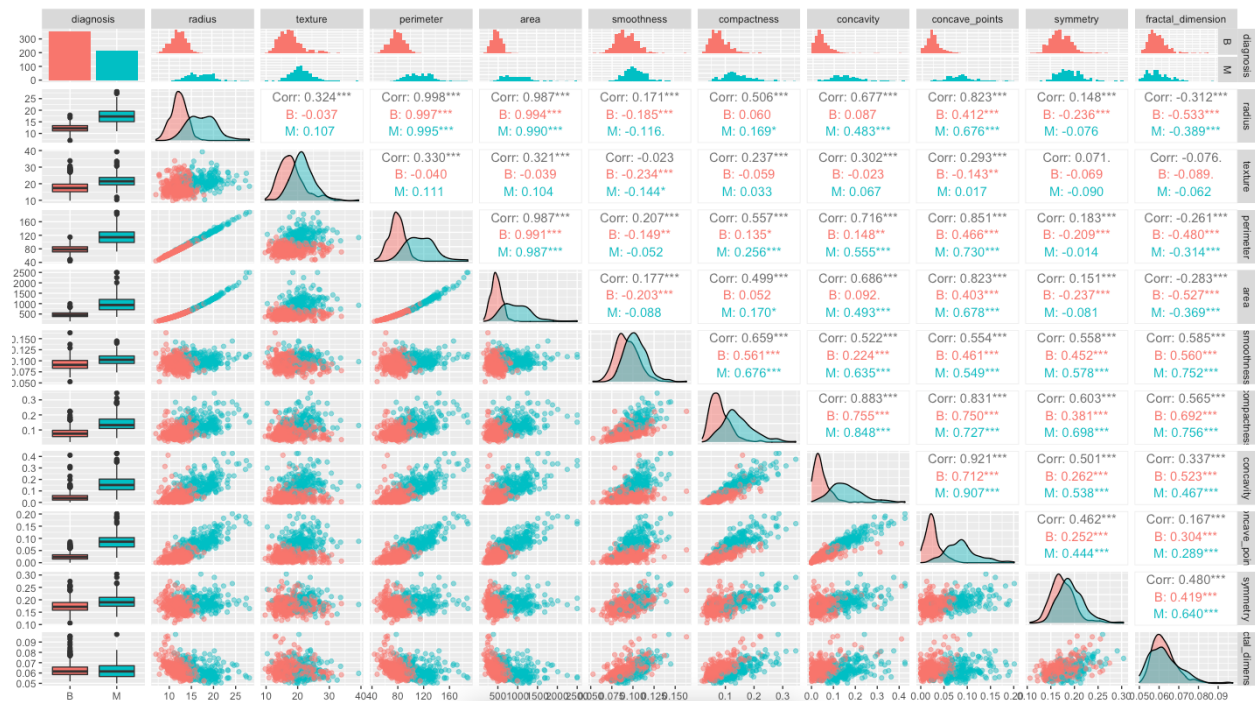


# Tutorial 10 Worksheet AY 22/23 Sem 1

DSA2101

In this tutorial, we continue with the Wisconsin Breast Cancer Dataset from Tutorial 8. The following plot is created using the GGally package. It shows scatterplots of all combinations of continuous variables in the dataset.



Take some time to study the plots. Remember that, for this dataset, we are interested in which variables would aid in predicting the diagnosis variable.

What observations can you make?

## Questions

In each of the questions below, spend some time after making each plot to consider what's good/bad about it, and then try to improve it. Take this opportunity to practice the iteration over charts to improve the visualisation.

1. Pick at least two continuous variables along with **diagnosis**, and create a chart of your own.
2. Create a parallel coordinates plot of the dataset, just like in the lecture notes (topic 05). How can you improve it to flesh out the relationship between the variables and **diagnosis**?